

**SECOND OPERATIONAL TEST AND EVALUATION
OF FORMS 18 AND 19 OF THE
ARMED SERVICES VOCATIONAL APTITUDE BATTERY**

Discussion and References

**Gary L. Thomasson, Lauress Wise,
Bruce Bloxom, John Welsh
*Defense Manpower Data Center***

Richard Branch
Military Entrance Processing Command

FEBRUARY 1995

**SECOND OPERATIONAL TEST AND EVALUATION
OF FORMS 18 AND 19 OF THE
ARMED SERVICES VOCATIONAL APTITUDE BATTERY**

Gary L. Thomasson, Lauress Wise

Bruce Bloxom, John Welsh

Defense Manpower Data Center

Richard Branch

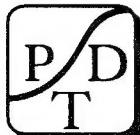
Military Entrance Processing Command

19950310 101



FEBRUARY 1995

Approved for public release; distribution is unlimited.



Personnel Testing Division
DEFENSE MANPOWER DATA CENTER



NOTE

This report, covering the **SECOND OPERATIONAL TEST AND EVALUATION OF FORMS 18 AND 19 OF THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY**, has been produced in two sections to facilitate review.

The front section contains the text that discusses the procedures and analyses and lists references.

The second section, titled the **ASVAB 18/19 SOT&E SUPPLEMENT**, contains the tables and figures that provide information to support the discussion of the procedures and analyses.

This report was prepared for the Directorate of Accession Policy, Office of the Assistant Secretary of Defense for Force Management Policy. The technical project officer for this report was Dr. Gary L. Thomasson, Quality Control and Analysis Branch, Personnel Testing Division, Defense Manpower Data Center, Monterey, CA. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official Department of Defense position, policy, or decision, unless so designated by other official documentation.

**SECOND OPERATIONAL TEST AND EVALUATION
OF FORMS 18 AND 19 OF THE
ARMED SERVICES VOCATIONAL APTITUDE BATTERY**

Discussion and References

**Gary L. Thomasson, Lauress Wise,
Bruce Bloxom, John Welsh
*Defense Manpower Data Center***

**Richard Branch
*Military Entrance Processing Command***

FEBRUARY 1995

**SECOND OPERATIONAL TEST AND EVALUATION
OF FORMS 18 AND 19 OF THE
ARMED SERVICES VOCATIONAL APTITUDE BATTERY**

Discussion and References

**Gary L. Thomasson, Lauress Wise,
Bruce Bloxom, John Welsh
*Defense Manpower Data Center***

**Richard Branch
*Military Entrance Processing Command***

FEBRUARY 1995

CONTENTS

| | |
|--|----|
| INTRODUCTION | 1 |
| METHOD | 3 |
| Design | 3 |
| Subjects..... | 3 |
| Procedure | 3 |
| DATA QUALITY CONTROL..... | 4 |
| Construction of Data Files | 4 |
| Construction of Equivalent Groups | 4 |
| Further Data Quality Control and Editing | 6 |
| Testing Equivalence of Groups..... | 6 |
| Testing Equivalence of Groups for Grade-specific Re-equating and Analyses... | 7 |
| GRADE-SPECIFIC RE-EQUATING: A SUMMARY..... | 8 |
| The Need for Grade-specific Re-equating | 8 |
| The Uses of Sample One and Sample Two | 8 |
| Data Edits for Grade-specific Re-equating | 9 |
| Item Order Effects for Grade-specific Re-equating | 9 |
| Calibration of Test for Grade-specific Re-equating (Overview) | 10 |
| RESULTS | 10 |
| Comparisons in the IOT&E Metric by Grade Level..... | 10 |
| Comparisons in the Re-equated Grade-specific Metric (ANOVAs) | 12 |
| Comparisons in the Re-equated Grade-specific Metric (t-tests)..... | 12 |
| SUMMARY AND CONCLUSIONS | 13 |
| REFERENCES | 15 |

| | |
|--------------------|-------------------------------------|
| Accession For | |
| WES GRA&I | <input checked="" type="checkbox"/> |
| DTEC TAB | <input type="checkbox"/> |
| Unpublished | <input type="checkbox"/> |
| Justification | |
| By _____ | |
| Description | |
| Availability Codes | |
| 2400 | Avail and/or Special |
| A-1 | _____ |

PREFACE

The completion of this work would not have been possible without the efforts of many persons at the Defense Manpower Data Center (DMDC) and elsewhere: the Military Entrance Processing Command provided the data files; Mr. Robert Hamilton of DMDC meticulously developed and documented the data base for the analyses; and Ms. Gretchen Glick of DMDC contributed a disciplined editorial eye and careful attention to all the necessary details in the final editing and production of the report.

The authors wish to thank the members of the Defense Advisory Committee (DAC) for their review of earlier drafts of this document and their thoughtful and detailed comments. Among the DAC members, we especially wish to thank Dr. Michael Kolen and Dr. Mark Reckase of American College Testing Program, Dr. Rebecca Zwick of Educational Testing Service, and Dr. Kevin Murphy of Colorado State University, for their valuable comments and consultation.

SECOND OPERATIONAL TEST AND EVALUATION OF FORMS 18 AND 19 OF THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY

INTRODUCTION

The *Armed Services Vocational Aptitude Battery* (ASVAB) is a set of ten tests administered by the Department of Defense (DoD) to all applicants for enlistment in the United States Armed Services, as part of the Enlistment Testing Program, and also administered in high schools as a component of the Student Testing Program. The tests included in ASVAB are listed in Table 1, which also shows the length of each test in items and minutes and the means and standard deviations of the number-correct scores for the 1980 Youth Population (Department of Defense, 1982). The battery yields scores for each of the ten tests, plus an eleventh composite score, Verbal (VE), which is the sum of scores on two tests, Word Knowledge (WK) and Paragraph Comprehension (PC). Various other combinations of the test scores form composites that are used by DoD and the Services for determining overall qualification for enlistment and eligibility for entry into specific occupational specialties.

The Student Testing Program serves two purposes. First, it provides students with career-guidance information. Second, the program allows students to qualify for enlistment into the Armed Services. The ASVAB scores that a student obtains in the Student Testing Program can be used to qualify that student for enlistment for up to two years after the date of testing. DoD provides this program to give recruiters an opportunity to present information on military careers and to generate lists of students who, on the basis of their ASVAB scores, have already passed aptitude requirements for enlistment.

ASVAB Forms 18 and 19 were developed for the Student Testing Program to replace Form 14. Two versions of each form, containing the same items in different orders, were developed. These versions were initially designated 18a and 18b, and 19a and 19b. The initial versions were changed slightly to accommodate a new answer sheet format necessitated by a change in the Optical Mark Readers (OMRs) used in scoring the tests (Bloxom, et al. 1993). The revised versions are designated 18f and 18g, and 19f and 19g.

Because Student Testing Program scores can be used for enlistment, it is essential that the ASVAB forms used in the Student Testing Program be parallel to the forms used in the Enlistment Testing Program and that scores from these forms be reported on the same scales used to report enlistment testing scores. The procedures used to develop ASVAB Forms 18 and 19 were the same as the procedures used in the development of other forms of the ASVAB (Curran & Palmer, 1990), and two prior equating studies were conducted to calibrate the scores generated with ASVAB Forms 18 and 19 (Bloxom & McCully, 1992).

The basic scales used for the individual tests are "standardized test" scores. These scales were defined by standardizing the number-correct (raw) scores on the Reference Form to have a mean of 50 with a standard deviation of 10 in the 1980 Youth Population (Department of

Defense, 1982). The standardization involves a linear translation of the number-correct scores, rounding to the nearest integer, and truncation of scores to a minimum of 20 and a maximum of 80. New forms of the ASVAB are calibrated by developing conversion tables for each test that give the most appropriate standardized test score for each number-correct score. The procedures used in developing these conversion tables are the same as those used in several other equating studies at the Defense Manpower Data Center (DMDC), the executive agent for the ASVAB (e.g., Bloxom, Thomasson, Wise, & Branch, 1992; Thomasson, Bloxom, & Wise, 1994).

The primary purpose of the Secondary Operational Test and Evaluation (SOT&E) of ASVAB Forms 18 and 19 was to monitor the test scores for accuracy and for the possibility of compromise. Accuracy is assessed by checking that the distribution of scores does not vary systematically by form. Compromise is checked by looking for local differences between some or all of the new forms and the reference form that might indicate coaching on a specific form within that location. Monitoring the possibility of test form compromise was of particular importance for ASVAB Forms 18 and 19 because the sequence numbering and distribution of these forms in an initial printing was flawed. While there was no reason to suspect a compromise, it was not possible to be certain that all forms were accounted for. Thus, an early check for test form compromise was deemed essential. Again, since students can use their ASVAB scores for accession into the military for up to two years, it is important to determine if there are any local anomalies or form differences when the operational calibration is applied in local student populations. For this purpose, the primary concern was in the Armed Forces Qualifications Test (AFQT) composite (involving tests AR, MK, and VE) since the AFQT score is used to determine overall qualification for enlistment.

The second purpose of the SOT&E data collection, which will be addressed in a separate report, was to evaluate the operational calibration of the new forms in the student population. One consequence of using the equipercentile approach to equate observed scores from different forms is that, to the extent the observed scores contain measurement error, the results will depend on the particular population used in the equating study (see Braun & Holland, 1982; Thomasson, 1993). Measurement error leads to a "regression to the mean" effect which will vary for different populations with different means. To the extent that the forms are truly parallel, reliability is high, or the equating populations are similar, the effects of this population dependency will be minimal.

The initial calibration of ASVAB Forms 18 and 19 was conducted using applicants for enlistment. This ensured that scores would be on a scale equated to the reference form as precisely as possible in the applicant population and was essential since ASVAB Form 18 and 19 scores can be used for enlistment. The population of students who participate in the Student Testing Program may be somewhat different from the population of applicants. Thus, the second very important purpose of the SOT&E data collection was to evaluate the operational calibration of the new forms in the student population.

METHOD

Design

The design of this study was to administer five ASVAB forms to randomly equivalent groups with approximately equal numbers of examinees per ASVAB form. The five forms were ASVAB Forms 18f, 18g, 19f, 19g, plus ASVAB 18h. Except for its cover, ASVAB Form 18h was identical to ASVAB Form 8a, the reference form which was used to collect the normative data (Department of Defense, 1982; normative means and standard deviations presented in Table 1). Two sets of comparisons were made among the score distributions of the five ASVAB form groups. The first used the operational equating conversion tables developed in the Initial Operational Test and Evaluation (IOT&E) (Bloxom & McCully, 1992). The second used two new grade-specific re-equating conversion tables based on the current student data. The comparisons after re-equating were used to assess the possibility of compromise of the new test forms after an adjustment for the population dependence of equating.

Subjects

The subjects in this study were all high school students participating in the operational Student Testing Program during the period 1 July 1992 through 8 December 1992. These subjects took the ASVAB as a part of this program. About 900,000 students were tested during this period. The exact sample sizes used in each procedure are indicated below.

Procedure

The subjects were tested in groups (test sessions) that varied in size according to the number of students being tested at a test site on a particular test day. Test administrators were generally hired and managed by the Office of Personnel Management under a contractual arrangement with Headquarters, U.S. Military Entrance Processing Command (HQ USMEPCOM). In some instances MEPCOM personnel served as test administrators. Additional proctors, as required, were generally provided by the recruiting services.

Each subject was provided with an answer sheet (circular response format), an ASVAB test booklet, two pencils, and two pieces of scratch paper. To provide equivalent conditions and frequency of administration for the five test forms, the forms were to be distributed in a spiraled order such that each form was administered to every fifth subject in a test session. Furthermore, the cycle of distribution of forms in each session was to begin where it stopped in the test administrator's previous session.

Before administration of the ASVAB, subjects were given standard instructions (Department of Defense, 1990) for entering identifying information and for signing a Privacy Act statement that appears on the answer sheet. The tests were then administered as specified in the standard ASVAB instructions. Following the test administration, the answer sheets were scanned and scored at the MEPS. Some answer sheets were further processed by a contractor, Human

Resources Research Organization (HumRRO), where they were rescanned to obtain individual item response data for each test.

DATA QUALITY CONTROL

Construction of Data Files

Two tapes were received by DMDC that contained data from the Student Testing Program during the period July 1992 through November 1992. The first was the Student Testing Program file from USMEPCOM which contained 910,805 records (before data editing). A second file containing item-level data on 311,339 participants in the Student Testing Program (before data editing) was received from HumRRO. After matching on the unique Answer Sheet Sequence Number (Litho Code), a merged file containing 911,343 cases was created. This merged file contained 310,809 records in which the Answer Sheet Sequence Number matched records in both source files, 599,996 records which appeared only in the MEPCOM data file, and 538 which appeared only in the file from HumRRO. It was decided to limit the initial analysis to matched records from both files so that item-level and score-level results would be based on the same samples.

Thus, the total number of records available in this first sample (called Sample One) was the 310,809 that matched on the unique Answer Sheet Sequence Number. However, only 63 of the 68 Military Entrance Processing Station (MEPS) regions were represented in the Sample One data, so many analyses were repeated using a less restrictive sample, called Sample Two, which contained data from all 68 MEPS. Sample Two started with the 911,343 merged records, before data editing.

Construction of Equivalent Groups

During the data collection, the five test forms were to be distributed in a spiraled fashion to subjects in each testing session. This stratification of test form administration was intended to provide five randomly equivalent groups of subjects. However the strict spiraled administration was not accomplished at each test session. If not corrected, the effect of unbalanced test administration across test sites with different mean ability levels might not have been distinguishable from the effect of test compromise. To avoid the confounding of effects, some procedure was needed to extract a "balanced" subset of the data.

The procedure used here to ensure a balanced administration of forms (and thus construct randomly equivalent groups as much as possible) was conservative in the sense that all data from questionable sessions were excluded from the analysis. Entire test sessions were eliminated from further data analysis if they did not meet certain criteria. Test sessions were included in the data for further analysis only if either:

- (1) (a) the Pearson chi-square test that frequency differences across ASVAB forms was due to sampling variations was less than 2.0, and (b) no individual cell's contribution to this chi-square was greater than or equal to 0.8,

or

- (2) the difference between the most frequently administered ASVAB form and the least frequently administered ASVAB form was less than 3 for that test session.

Criterion (1) was used to discard imbalance in moderate to large test sessions, and criterion (2) was to tolerate small imbalances in small sized test sessions that might not pass the first criterion. The critical values in criterion (1) were decided on after visual inspection of numerous plots of the data.

The graphs shown in Figures 1a - 1d for Sample One and Figure 2a - 2d for Sample Two show the results of this editing procedure. Figure 1a plots the percentage of form administered in a "deleted" session versus the total sample size of the session for Sample One. Each session is represented by five data points, one for each of the five forms in this study. The target for a "balanced" session is 20% for each of the five forms. Note that one or more of the forms in a "deleted" session can be exactly 20%, but the entire session can still be deleted by failing the "balanced" criteria for the entire session. Figure 1b plots the percentage of form administered in a "kept" session, versus the total sample size of the session for Sample One. In this plot, note that the "balanced" criteria forces all points to fall near the target of 20% except for sessions with small sample sizes. Figure 1c plots the percentage difference between the form with the highest percentage administered in a "deleted" session and the form with the lowest percentage administered in the same session versus the total sample size of the session for Sample One. Here there is only one data point per session, and the target is a zero difference in the percentages. Figure 1d plots the same maximum percentage difference in form administrations as in Figure 1c except for only the "kept" or "balanced" sessions for Sample One. The descriptions for Figures 2a - 2d follow those of Figure 1a - 1d respectively, except they represent Sample Two results. Although such data editing for "balanced" sessions may appear to be severe, the remaining samples sizes were still large enough to provide statistically precise results.

For Sample One, there were a total of 7,007 test sessions in the matched data set of 310,809 examinees. A total of 1,469 test sessions were deleted which included 91,518 examinees. Thus, 219,291 examinees were available in the "balanced" Sample One data set for further analysis (Table 2). For Sample Two, there were a total of 12,858 test sessions in the merged data set of 911,343 examinees. A total of 3,553 test sessions were deleted which included 355,992 examinees. Thus, 555,351 examinees were available in the "balanced" Sample Two data set for further analysis (Table 3).

Further Data Quality Control and Editing

Table 2 and Table 3 show the distributions of surviving cases across test forms and the percentage of cases deleted as a result of each of the editing procedures for Sample One and Sample Two respectively. In addition to the editing for "balanced" groups, checks were made for a valid form code and for "below chance" responding.

The "below chance" editing procedure eliminated cases (individual examinees) with three or more test number-right scores at or below chance responding. Such cases might have resulted from low motivation. The decision to remove cases having three or more scores at or below chance is based on procedures and judgments developed at the Air Force Human Resources Laboratory (1988) and replicated by DMDC in the Optical Mark Reader (OMR) calibration study (Bloxom, et al., 1990). This "below chance" editing procedure resulted in the elimination of 14,049 of the 219,291 cases from Sample One and of 36,001 of the 555,351 cases from Sample Two. The last columns of Table 2 and of Table 3 show the distributions of surviving cases across test forms and the percentage of cases deleted as a result of the "below chance" editing procedures for Sample One and Sample Two respectively.

Testing Equivalence of Groups

As a partial check on the equivalence-of-groups across sites, a statistical test was made of equal proportions of ASVAB forms across MEPS sites. For analyses reported here, each MEPS with its associated METs was considered a MEPS site. For the 219,291 cases in the Sample One "balanced" edited data set, the test for equal proportions of ASVAB Forms across the 63 represented MEPS sites was not significant ($\chi^2 = 34.312$, d.f. = 248, $p > 0.05$). For the 555,351 cases in the Sample Two "balanced" edited data set, the test for equal proportions of ASVAB forms across the 68 represented MEPS sites was not significant ($\chi^2 = 45.731$, d.f. = 272, $p > 0.05$).

In addition to checks on the "spiraling" procedures, more direct checks of group equivalence were performed. If the five test-form groups differed on characteristics that are typically correlated with test performance, using the data for equipercentile equating would require adjustments of the distributions. Therefore, as a check on group equivalence, the five test-form groups were compared with respect to three background characteristics -- gender, race and education -- that were reported by the examinees on their answer sheets. Table 4a, b, c, provides frequencies and percentages at each level of these variables for each of the five test-form groups. The group-by-test-form Pearson chi-squares were not statistically significant ($p > .05$) for any of the three background characteristics using Sample One "balanced" data. Since all checks for group non-equivalency (equivalence of form groups across MEPS sites, and background characteristics) were non-significant for Sample One, the groups were considered sufficiently equivalent to justify proceeding with further analyses of that sample.

The group-by-test-form Pearson chi-squares for the same three background characteristics using Sample Two "balanced" data are given in Table 5a, b, c. Except for Education Level, the chi-squares are not significant at alpha = 0.05. However, since the significance was only marginal, the chi-square for Education Level by form was less than twice the degrees of freedom, and several tests were being made, the groups in the Sample Two data were also considered sufficiently equivalent for further analyses.

Testing Equivalence of Groups for Grade-specific Re-equating and Analyses

As described below, we were concerned with grade-specific re-equating and grade-specific analyses for grades 12 and 11. Thus we were concerned with checks on the equivalence of groups for each of these grade levels separately, in addition to the combined grades. As a partial check on the equivalence-of-groups for within grade level, a statistical test was made of equal proportions of forms across 63 represented MEPS sites in the Sample One data. For the 58,597 Grade 12 students in the "balanced" Sample One edited data set, the test for equal proportions of forms across MEPS sites was not significant ($\chi^2 = 97.944$, d.f. = 248, $p > 0.05$). Likewise for the 120,218 Grade 11 students in the "balanced" Sample One edited data set, the test for equal proportions of forms across MEPS sites was not significant ($\chi^2 = 74.350$, d.f. = 248, $p > 0.05$).

The other checks on group equivalence were the chi-square tests for independence of each of the two background characteristics of gender and race across the five test-form groups. Table 6a, b, c and Table 7a, b, c provide frequencies and percentages at each level of these variables for each of the five test-form groups for Grade 12 subjects and Grade 11 subjects respectively for Sample One data. The group-by-test-form Pearson chi-squares were not statistically significant ($p > .05$) for any of the three background characteristics. Since all checks for group non-equivalency (equivalence of form groups across MEPS sites and background characteristics) were non-significant, the groups were considered sufficiently equivalent to justify proceeding with the test form equating computations for Grade 12 and Grade 11 using their respective Sample One data sets.

As a partial check on the equivalence-of-groups for each grade level separately, statistical tests were made for equal proportions of five ASVAB Forms across the 68 represented MEPS sites for each grade level in Sample Two. For the 155,515 Grade 12 students in the "balanced" Sample Two edited data set, the test for equal proportions of forms across MEPS sites was not significant ($\chi^2 = 179.510$, d.f. = 268, $p > 0.05$). Likewise for the 304,527 Grade 11 students in the "balanced" Sample Two edited data set, the test for equal proportions of forms across MEPS sites was not significant ($\chi^2 = 122.716$, d.f. = 268, $p > 0.05$).

The other checks on group equivalence were the chi-square tests for independence of each of the two background characteristics of gender and race across the five test-form groups. Table 8a, b, c provides frequencies and percentages at each level of these variables for each of the five test-form groups for Grade 12 Sample Two subjects. Table 9a, b, c provides frequencies and

percentages at each level of these variables for each of the five test-form groups for Grade 11 Sample Two subjects. The group-by-test-form Pearson chi-squares were not statistically significant ($p > .05$) for any of the background characteristics except for the race variable. Even in this case, the chi-square was less than twice the degrees-of-freedom and judged to be small enough to consider groups roughly equivalent. Thus, the groups were considered sufficiently equivalent to justify proceeding with further analyses of the 12th and 11th grade Sample Two edited data sets.

GRADE-SPECIFIC RE-EQUATING: A SUMMARY

The Need for Grade-specific Re-equating

As previously stated, one consequence of using the equipercentile approach to equate observed scores from different forms is that, to the extent that the alternate forms are not strictly parallel and that the observed scores contain measurement error, the results depend on the particular population used in the equating study. Measurement error leads to a "regression to the mean" effect which will vary for different populations with different means. To the extent that the forms are truly parallel, reliability is high, or the equating populations are similar, the effects of this population dependency will be minimal.

If there are population dependency effects between the IOT&E applicant population used to compute the equating and the SOT&E students participating in this study, the analyses will be contaminated by population dependency effects. To eliminate the effect of this population dependency of equating functions from the SOT&E analyses, new equatings were computed separately for the 12th and 11th grades. Grades 12 and 11 were selected for analyses because these are the grades for which the Student AFQT scores could be used for accession by recent high school graduates because of the two year limit on the use of scores for enlistment, and these grade levels had the substantial numbers needed for the grade level analyses (Tables 4a, b, c, and 5a, b, c). Grade-specific re-equatings were performed since it seemed that Grade 12 and Grade 11 populations were reliably different from each other, as well as from the applicant population.

The Uses of Sample One and Sample Two

The tests based on the re-equated scores assume that any compromise is locally limited rather than nationwide or broadly dispersed. Sample One, a well "balanced" sample for which item level data exist, was used for computing a new equating function. Sample One was also comparable in sample size (more than 10,000 per form) to typical recent DMDC IOT&E equating studies (e.g., Thomasson, Bloxom, & Wise, 1994). To test for compromise, Sample Two, the larger but less well "balanced" sample, was used in these analyses. The obvious reasons for using Sample Two data to investigate compromise were the increased sample size and statistical power, as well as the fact that five MEPS were completely missing in the Sample One data.

Data Edits for Grade-specific Re-equating

The data used in the grade-specific re-equating started with subjects who reported their education level as Grade 12 or Grade 11, respectively, from the edited file for the full Sample One data set. An additional data edit step was made when computing the distributions of raw test scores for recalibration on a variable-by-variable level rather than a case-by-case level. That is, all tests with scores of zero were removed from the distributions of raw scores just before they were used to calibrate the equating transformations. However, these scores were not eliminated from the data sets for other subsequent analyses, e.g., for tests of mean differences. Table 10 contains statistics and sample sizes for the distributions from Sample One used for the Grade 12 re-equating, while Table 11 contains the same information from Sample One Grade 11 distributions used for re-equating.

Item Order Effects for Grade-specific Re-equating

Each test in each of the ASVAB Forms 18 and 19 contains the same items that are in one of the other forms in the study.

- For each test that contributes to the AFQT composite (AR, WK, PC, MK, and VE), Forms 18f and 19f contain the same items, and Forms 18g and 19g contain the same items.
- For each of the remaining tests (GS, NO, CS, AS, MC, and EI), Forms 18f and 18g contain the same items, and Forms 19f and 19g contain the same items.

For every test, with the exception of NO, the items differ slightly in the order of administration in the two forms that contain them. The purpose of this slight scrambling of item order was to make it unlikely that a subject could obtain correct answers by copying responses from the answer sheet of another subject who was administered a different ASVAB form.

Two forms of a test with the same items in a slightly different order may not have the same distribution of scores and may, therefore, require separate equatings to the reference form. A statistical procedure was used to assess item-order effects for each test on each pair of forms containing scrambled orderings of the same items. If the statistical test was found to be significant for a pair of same-item forms, then separate equatings would be done for each separate form. Otherwise, the data for the same-item forms could be combined and a single equating developed for use with either form. The results are summarized in Table 12 for Grade 12 and in Table 13 for Grade 11. The details of the procedures for the statistical tests for item order effects are the same as those used in several other equating studies at DMDC (Bloxom, Thomasson, Wise, & Branch, 1992; Thomasson, Bloxom, & Wise, 1994).

Calibration of Tests for Grade-specific Re-equating (Overview)

The selected method of equating each of the new forms to the reference form was that of equipercentile equating for equivalent groups using a *polynomial log-linear* distribution smoothing procedure. The procedures used for calibration of tests for grade-specific re-equating are same as those used in several other equating studies at DMDC. For more details of the procedures used in the calibration of tests, see Bloxom, Thomasson, Wise, & Branch (1992) and Thomasson, Bloxom, & Wise (1994).

RESULTS

The scores for the Grade 12 students and the Grade 11 students were separately analyzed in two metrics: (1) in the original IOT&E metric for reporting scores of record should the student enlist on his/her student ASVAB, and (2) in the re-equated metric using the grade-specific re-equatings. As reported in an earlier section, statistical checks were made of the equivalence-of-groups assumption for each grade level separately in Sample Two.

Comparisons in the IOT&E Metric by Grade Level

The standard scores that are reported for the Student Testing Program are based on test equating and calibration using data collected in Initial Operational Test and Evaluation (IOT&E) studies. The calibrations for the IOT&E of ASVAB Forms 18 and 19 were based on the population of applicants for enlistment in the military that were processed during January and February 1991 (Bloxom & McCully, 1992). This population of applicants would almost assuredly differ from the population of Grade 12 or Grade 11 students in the present study in terms of variables relevant to test performance (i.e., the distribution of abilities, the motivational levels, etc.). However, even though equipercentile equatings are generally population dependent, there are cases in which such equatings may be relatively robust under some shifts in the population distribution of ability (Thomasson, 1993). Therefore, as a first step in looking for aberrant mean differences in measured ability across forms, the reported test standard scores based on the IOT&E equatings were examined.

As stated earlier, the scores of primary interest are those involved with the AFQT. The variable labeled STD-AFQT is the Sum of Subtest Standard Scores (SSSS) used to compute the AFQT such that $STD\text{-}AFQT = 2*VE + AR + MK$, where VE, AR, and MK are in standard score form (either using the IOT&E metric or the re-equated metric). The difference between the STD-AFQT score and the final computed AFQT score is a nonlinear transformation that takes the STD-AFQT score to a percentile scale based on the 1980 youth population norming sample. The

advantages of using STD-AFQT rather than the AFQT on a percentile scale is that STD-AFQT should be more "bell-shaped" or normally distributed, while the AFQT will tend to be more rectangularly distributed (at least in a population similar to the 1980 youth population). Using STD-AFQT thus makes the parametric statistical tests such as Analysis of Variance (ANOVA) and Student's t-test statistics more meaningful since these statistical tests are based on normal distribution theory.

A comparison in the IOT&E metric was made using an Analysis of Variance (ANOVA) of the IOT&E standard scores using Sample Two data. The ANOVA design was a MEPS-by-ASVAB test-form design. Since it was expected that the MEPS data would vary, the main "MEPS effect" was expected to be significant. And to the extent that the IOT&E equatings are population dependent, the main effect of the test form may be expected to be significant. However, a significant interaction between MEPS and the test form could imply that the test forms were functioning differently at different MEPS. A significant MEPS-by-test-form interaction could be accounted for by the compromise of one or more forms at one or more MEPS regions. However, other factors (e.g., school curricula differences) could also account for a significant interaction. More detailed analysis would be necessary to discover which if any forms might be compromised and at which MEPS.

Table 14a shows the summary of results from ANOVA on the Grade 12 Sample Two data set for the STD-AFQT scores. Table 15a shows the summary of results from ANOVA on the Grade 11 Sample Two data set STD-AFQT scores. For both grade levels, the MEPS site factor was significant at alpha = 0.05 as expected because of the varying ability levels of the populations taking the ASVAB tests across each MEPS region. The main effect of the ASVAB forms factor was also significant at alpha = 0.05 despite the efforts to ensure equivalent groups across forms within each tests session. With large sample sizes such as those used here, even small differences will register as statistically significant. As a more useful measure of relative sizes of these effects, a percentage of a "corrected" total model variance¹ was also computed for the main effects and interaction effects in each ANOVA. These proportions of total model variances, which allow us to compare the relative sizes of the separate effects, demonstrate that differences in MEPS accounts for most of the modeled effects. The differences in ASVAB forms accounts for a very small percentage of the total variance accounted for by the full model. Tables 14b and 15b contain the overall least-squares means for the ASVAB FORMS groups within each grade level for comparison. In the IOT&E metric, the means for Forms 18g and 19g are similar to the reference form mean while the means on Forms 18f and 19f are higher by one or two STD-AFQT score points. One reason for the unequal means may be due to the population dependency of the IOT&E equipercentile equating. The population of the Grade 12 and Grade 11 students participating in the Student Testing Program may be reliably different from the applicant population in the Enlistment Testing Program in terms of factors related to test

¹ The hypothesis tests and related statistics reproted here are based on the SAS Type III hypothesis tests. For the Type III tests, the Sum of Squares (SS) for each effect is computed so that it is indepenedent of other effects tested, and thus the simple sum of the Sum of Squares for all the effects does not normally add up to the full model SS. Thus, the "corrected" total model variance used in the computations here as a reference is simply the sum of the Sum of Squares for the individual effects -- and is less than the SS for the full model.

performance, such as ability levels at the time of testing, motivation levels at the time of testing, and prior practice or experience with the content of the ASVAB tests. Therefore one must look at comparisons using the population-appropriate equating function (i.e., the re-equated metric) for proper comparisons.

The interaction effects were also significant at alpha = 0.05 for both grade levels. If the interaction had been non-significant, this would have been partial evidence supporting an argument of no substantial test compromise. However, since many hypotheses could account for a significant interaction in addition to the compromise hypothesis, other analyses must be considered to argue for the compromise hypothesis. The alternative hypotheses include difference in curricula across MEPS regions and/or an interaction involving the population dependent equating transformations.

Comparisons in the Re-equated Grade-specific Metric (ANOVAs)

Table 16a shows the summary of results from ANOVA on the Grade 12 Sample Two data set for the STD-AFQT scores in the Grade 12 re-equated metric. Table 17a shows the summary of results from ANOVA on the Grade 11 Sample Two data set STD-AFQT scores in the Grade 11 re-equated metric. As in the IOT&E metric, the MEPS site factor was significant at alpha = 0.05 for both grade levels as expected because of the varying ability levels of the populations taking the ASVAB tests across each MEPS region. The main effect of the ASVAB forms factor was significant at alpha = 0.05 for Grade 12 but was not significant for Grade 11. (Tables 16b and 17b contain the overall least-squares means for the ASVAB forms groups within each grade level for comparison.) The interaction effects were also significant at alpha = 0.05 for both grade levels and were similar in magnitude as they were in the IOT&E metric analyses. From these ANOVA results more detailed analyses such as t-tests are indicated.

Comparisons in the Re-equated Grade-specific Metric (t-tests)

As follow-up to the ANOVA analyses in the re-equated metrics, the separate t-test analyses were performed comparing the mean re-equated STD-AFQT score for each ASVAB 18 and 19 form with that of the reference form separately for each MEPS region. The histogram for the distribution of t-test statistics for ASVAB Form 18f using Sample Two data are presented in Figure 3 for Grade 12 and in Figure 4 for Grade 11. Figures 5 and 6 are the corresponding t-test distributional histograms for ASVAB Form 19f for Grades 12 and 11 respectively. Figures 7 and 8 are the same for ASVAB Form 18g, and Figures 9 and 10 are the same for ASVAB Form 19g. These figures plot the frequency of occurrence of t-statistics in bins that are 0.5 t-units wide. For instance, the bin labeled 0 includes $-0.25 \leq t < 0.25$, and the bin labeled 0.5 includes $0.25 \leq t < 0.75$, etc. For comparison purposes, the normal distribution is indicated as the "expected distribution" in each graph. This normal distribution is a reasonable comparison for these statistics since the degrees-of-freedom for the t-statistics ranged from 162 to 3,174 for Grade 12 and from 91 to 5,484 for Grade 11.

Since Grade 12 and Grade 11 represent independent samples taking the same test forms, a bivariate plot of the t-statistics for Grade 12 versus Grade 11 should have a bivariate near-normal distribution with correlation of zero. Figure 11 shows the bivariate plot of Grade 12 versus Grade 11 for t-statistics on re-equated STD-AFQT scores for Form 18f using Sample Two data. Figures 12, 13, and 14 show the same bivariate plots for Forms 19f, 18g, and 19g respectively. Also remember that 19f is something like a replication of 18f, and 19g is something like a replication of 18g. Points on these bivariate plots are identified on all such plots by an identifying code number (that is not the usual identifier for MEPS) if that corresponding MEPS has any of the t-test statistics > 3.0 for any Grade level or ASVAB Form. While no MEPS has a t-value > 3.0 for both the 11th and 12th grades, several MEPS do have significant t-values for some forms in the 12th grade sample.

(Note that for this public report the identity of specific MEPS is not being released. However, all information is made available to the Manpower Accession Policy Working Group (MAPWG) and to HQ MEPCOM in separate documents.)

SUMMARY AND CONCLUSIONS

The primary purpose of this Secondary Operational Test and Evaluation of ASVAB Forms 18 and 19 in the Student Testing Program was to monitor scores for ASVAB form comparability across testing regions and assess the possibility of regional compromise of one or more test forms. The design of this study was to administer five ASVAB forms (Forms 18f, 18g, 19f, 19g, and 18h = reference form) to randomly equivalent groups in the operational Student Testing Program with approximately equal numbers of examinees per ASVAB form. Data editing procedures were required so that the five ASVAB form groups within any test session would be nearly equivalent in size. A preliminary Analysis of Variance in the operational standardized metric (the IOT&E metric) was made to assess mean form differences in STD-AFQT scores ($STD-AFQT = 2*VE + AR + MK$ in standard score form) across the different MEPS regions. The results of these analyses seemed to support the hypothesis that the proper metric for the study of compromise in the Student Testing Program involves grade-specific re-equating to avoid the confounding of population dependency of equating functions. ANOVAs of mean differences in STD-AFQT scores in the grade-specific re-equating metric suggested that more detailed analyses at the individual MEPS level were needed. T-tests for mean differences of each ASVAB form (18f, 18g, 19f, 19g) with the reference form (18h) were separately computed for each MEPS region. In general, the distributions of these t-test statistics roughly followed an expected normal distribution. The more extreme cases, while possibly occurring by chance, were marked for future follow-up study.

From the results examined thus far, there has been little evidence of any substantial local form differences or of substantial compromise of forms within particular MEPS testing regions.

Even though there is little evidence of substantial compromise, and the observed results may be due to chance alone, there are some indications that some MEPS regions with the more discrepant results deserve further scrutiny. Further and more detailed analyses with attention on the scores of separate forms, separate tests, and maybe even separate items at some individual MEPS may be indicated. Any additional findings will be passed along to USMEPCOM, as well as the Manpower Accession Policy Working Group, with the possible recommendation that follow-up checks be made at certain sites.

Among the possible additional analyses being considered are the following:

- (1) Modeling different types of test compromise and testing the models. For example, (a) testing for compromise of a test form versus compromise of single items, (b) modeling population ability differences and test form reliability (or measurement precision) differences, and (c) the use of appropriateness analyses in testing models of cheating (Drasgow, Levine, Williams, McCusker, Thomasson, & Lim, 1989; Levine & Drasgow, 1988).
- (2) Item level analyses including a Differential Item Functioning (DIF) type of analyses in which the focal group would be a suspect MEPS region.
- (3) Investigating sessions within the MEPS that are problematic to determine if mean form differences are due to a few sessions out of many or due to more general differences over all or in most sessions within a MEPS site. Trends within MEPS and Mobile Examining Team sites (METs) could also be analyzed.
- (4) Recommendations and procedures for continued score monitoring in the Student Testing Program (without the use of the reference form).

A secondary purpose of the SOT&E investigation was to evaluate the initial operational (IOT&E) calibration of the new forms in the student population. Comparisons of equating functions (IOT&E operational equating, Grade 12 re-equating, and Grade 11 re-equating) for each of the ASVAB forms in this study will be made in a subsequent equating-comparisons report. This follow-up report will investigate the extent of the differences in equating functions associated with different equating populations (samples) including the IOT&E applicant sample, the separate 11th and 12th grade samples analyzed in the present Student Testing Program study, and possibly the recruit sample used in the Operational Calibration equating study. Questions relevant to the appropriate use of various equating functions will be addressed in this future equating-comparisons report.

REFERENCES

- Air Force Human Resources Laboratory. (1988, May). *Scanning and data editing of ASVABs 18 and 19*. Unpublished paper, Brooks Air Force Base, TX.
- Bloxom, B., & McCully, R. (1992, August). *Initial operational test and evaluation of forms 18 and 19 of the Armed Services Vocational Aptitude Battery* (DMDC TR 92-001). Monterey, CA: Defense Manpower Data Center.
- Bloxom, B., McCully, R., Branch, R., Waters, B.K., Barnes, J., & Gribben, M. (1991, August). *Operational calibration of new optical-mark-reader answer sheets for the Armed Services Vocational Aptitude Battery*. Unpublished manuscript, Defense Manpower Data Center, Monterey, CA.
- Bloxom, B., McCully, R.A., Branch, R.L., Waters, B.K., Barnes, J., & Gribben, M. (1993, July). *Operational calibration of the circular-response optical-mark-reader answer sheets for the Armed Services Vocational Aptitude Battery (ASVAB)* (DMDC TR 93-009). Monterey, CA: Defense Manpower Data Center.
- Braun, H.I., & Holland, P.W. (1982). Observed test-score equating: a mathematical analysis of some ETS equating procedures. In P.W. Holland & D.B. Rubin (Eds.), *Test equating* (pp. 9-50). New York: Academic Press.
- Curran, L.T., & Palmer, P. (1990, February). *Armed Service Vocational Aptitude Battery (ASVAB): Item overlength and operational length development of forms 18 and 19* (AFHRL-TP-89-74). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.
- Drasgow, F., Levine, M.V., Williams, B., McCusker, C., Thomasson, G.L., & Lim, R.G. (1990, January). *Evaluation of optimal appropriateness measurement for use in practical settings* (AFHRL-TP-89-41). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.
- Haberman, S.J. (1974). Log-linear models for frequency tables with ordered classifications. *Biometrics*, 36, 589-600.
- Hanson, B.A. (1990a). *An investigation of methods for improving test score distributions* (ACT Report No. 90-4). Iowa City, IA: American College Testing Program.
- Hanson, B.A. (1990b). *Description of a program for smoothing univariate test score distributions*. Unpublished manuscript, American College Testing Program, Iowa City, IA.
- Hanson, B.A. (1992, April). *Testing for differences in test score distributions using log-linear models*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Holland, P.W., & Thayer, D.T. (1987). *Notes on the use of log-linear models for fitting discrete probability distributions* (TR 87-79). Princeton, NJ: Educational Testing Service.
- Kolen, M.J. (1991). Smoothing methods for estimating test score distributions. *Journal of Educational Measurement*, 28, 257-282.
- Levine, M.V., & Drasgow, F. (1988). Optimal appropriateness measurement. *Psychometrika*, 53, 161-176.

- Thomasson, G.L. (1993, August). *The asymptotic equating methodology and other test equating evaluation procedures*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Thomasson, G., & Bloxom, B. (1992, March). *Operational calibration of forms 20, 21, and 22 of the Armed Services Vocational Aptitude Battery*. Unpublished manuscript, Defense Manpower Data Center, Monterey, CA.
- Thomasson, G.L., Bloxom, B., & Wise, L. (1994, January). *Initial operational test and evaluation of forms 20, 21, and 22 of the Armed Services Vocational Aptitude Battery (ASVAB)* (DMDC TR 94-001). Monterey, CA: Defense Manpower Data Center.
- U.S. Department of Defense. (1982). *Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery*. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics).
- U.S. Department of Defense. (1990, January). *Manual for administration: Armed Services Vocational Aptitude Battery* (DoD 1304.12-L-AM-1). Chicago, IL: U.S. Military Entrance Processing Command.

ASVAB 18/19 SOT&E SUPPLEMENT

Tables 1-17 and Figures 1-14

**Gary L. Thomasson, Lauress Wise,
Bruce Bloxom, John Welsh
*Defense Manpower Data Center***

Richard Branch
Military Entrance Processing Command

FEBRUARY 1995

CONTENTS

TABLES

| | | |
|----------|--|------|
| Table 1 | ASVAB Tests, Number of Items, Time Limits, Normative Means, and Standard Deviations | S-1 |
| Table 2 | Sample 1: Number and Percentages of Subjects by Test Form and Edit Procedure..... | S-2 |
| Table 3 | Sample 2: Number and Percentages of Subjects by Test Form and Edit Procedure..... | S-3 |
| Table 4a | Sample 1: Number and Percentages of Subjects by Gender | S-4 |
| Table 4b | Sample 1: Number and Percentages of Subjects by Education..... | S-5 |
| Table 4c | Sample 1: Number and Percentages of Subjects by Race..... | S-6 |
| Table 5a | Sample 2: Number and Percentages of Subjects by Gender | S-7 |
| Table 5b | Sample 2: Number and Percentages of Subjects by Education..... | S-8 |
| Table 5c | Sample 2: Number and Percentages of Subjects by Race..... | S-9 |
| Table 6a | Sample 1: Grade 12 Number and Percentages of Subjects by Gender | S-10 |
| Table 6b | Sample 1: Grade 12 Number and Percentages of Subjects by Population Group | S-11 |
| Table 6c | Sample 1: Grade 12 Number and Percentages of Subjects by Post-High School Intention | S-12 |
| Table 7a | Sample 1: Grade 11 Number and Percentages of Subjects by Gender | S-13 |

| | | |
|-----------|--|------|
| Table 7b | Sample 1: Grade 11 Number and Percentages of Subjects by Population Group | S-14 |
| Table 7c | Sample 1: Grade 11 Number and Percentages of Subjects by Post-High School Intention | S-15 |
| Table 8a | Sample 2: Grade 12 Number and Percentages of Subjects by Gender | S-16 |
| Table 8b | Sample 2: Grade 12 Number and Percentages of Subjects by Population Group | S-17 |
| Table 8c | Sample 2: Grade 12 Number and Percentages of Subjects by Post-High School Intention | S-18 |
| Table 9a | Sample 2: Grade 11 Number and Percentages of Subjects by Gender | S-19 |
| Table 9b | Sample 2: Grade 11 Number and Percentages of Subjects by Population Group | S-20 |
| Table 9c | Sample 2: Grade 11 Number and Percentages of Subjects by Post-High School Intention | S-21 |
| Table 10 | Sample 1: Grade 12 Test Raw Score Mean, Standard Deviation, Skewness, Kurtosis, and Number of Subjects by Form | S-22 |
| Table 11 | Sample 1: Grade 11 Test Raw Score Mean, Standard Deviation, Skewness, Kurtosis, and Number of Subjects by Form | S-25 |
| Table 12 | Sample 1: Grade 12 Likelihood Ratio Chi-Square Tests of Significance Item Order Effects..... | S-28 |
| Table 13 | Sample 1: Grade 11 Likelihood Ratio Chi-Square Tests of Significance Item Order Effects..... | S-29 |
| Table 14a | Sample 2: Grade 12 STD-AFQT IOT&E METRIC ANOVA..... | S-30 |

| | |
|--|------|
| Table 14b Sample 2: Grade 12 | |
| STD-AFQT IOT&E METRIC ANOVA | |
| LEAST SQUARES MEANS..... | S-30 |
| Table 15a Sample 2: Grade 11 | |
| STD-AFQT IOT&E METRIC ANOVA..... | S-31 |
| Table 15b Sample 2: Grade 11 | |
| STD-AFQT IOT&E METRIC ANOVA | |
| LEAST SQUARES MEANS..... | S-31 |
| Table 16a Sample 2: Grade 12 | |
| STD-AFQT RE-EQUATED METRIC ANOVA | S-32 |
| Table 16b Sample 2: Grade 12 | |
| STD-AFQT RE-EQUATED METRIC ANOVA | |
| LEAST SQUARES MEANS..... | S-32 |
| Table 17a Sample 2: Grade 11 | |
| STD-AFQT RE-EQUATED METRIC ANOVA | S-33 |
| Table 17b Sample 2: Grade 11 | |
| STD-AFQT RE-EQUATED METRIC ANOVA | |
| LEAST SQUARES MEANS..... | S-33 |

FIGURES

| | |
|--|------|
| Figure 1a Percentage of Form (18F, 18G, 19F, 19G, & Reference) Administered in a “Deleted” Session by Total Number of Forms Administered Per Session for Sample #1 | S-34 |
| Figure 1b Percentage of Form (18F, 18G, 19F, 19G, & Reference) Administered in a “Kept” Session by Total Number of Forms Administered Per Session for Sample #1 | S-35 |
| Figure 1c Percentage Difference (Most Administered-Least Administered Form) in a “Deleted” Session by Total Number of Forms Administered Per Session for Sample #1 | S-36 |
| Figure 1d Percentage Difference (Most Administered-Least Administered Form) in a “Kept” Session by Total Number of Forms Administered Per Session for Sample #1 | S-37 |
| Figure 2a Percentage of Form (18F, 18G, 19F, 19G, & Reference) Administered in a “Deleted” Session by Total Number of Forms Administered Per Session for Sample #2 | S-38 |
| Figure 2b Percentage of Form (18F, 18G, 19F, 19G, & Reference) Administered in a “Kept” Session by Total Number of Forms Administered Per Session for Sample #2 | S-39 |
| Figure 2c Percentage Difference (Most Administered-Least Administered Form) in a “Deleted” Session by Total Number of Forms Administered Per Session for Sample #2 | S-40 |
| Figure 2d Percentage Difference (Most Administered-Least Administered Form) in a “Kept” Session by Total Number of Forms Administered Per Session for Sample #2 | S-41 |
| Figure 3 Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 18F versus Reference Form: Grade 12; Sample #2..... | S-42 |
| Figure 4 Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 18F versus Reference Form: Grade 11; Sample #2..... | S-42 |
| Figure 5 Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 19F versus Reference Form: Grade 12; Sample #2..... | S-43 |

| | | |
|-----------|---|------|
| Figure 6 | Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 19F versus Reference Form: Grade 11; Sample #2..... | S-43 |
| Figure 7 | Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 18G versus Reference Form: Grade 12; Sample #2..... | S-44 |
| Figure 8 | Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 18G versus Reference Form: Grade 11; Sample #2..... | S-44 |
| Figure 9 | Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 19G versus Reference Form: Grade 12; Sample #2..... | S-45 |
| Figure 10 | Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 19G versus Reference Form: Grade 11; Sample #2..... | S-45 |
| Figure 11 | Bivariate Plot of Grade 12 by Grade 11 T-Test Statistics for Mean Difference of STD-AFQT Score: Form 18F versus Reference Form: Sample #2 | S-46 |
| Figure 12 | Bivariate Plot of Grade 12 by Grade 11 T-Test Statistics for Mean Difference of STD-AFQT Score: Form 19F versus Reference Form: Sample #2 | S-46 |
| Figure 13 | Bivariate Plot of Grade 12 by Grade 11 T-Test Statistics for Mean Difference of STD-AFQT Score: Form 18G versus Reference Form: Sample #2..... | S-47 |
| Figure 14 | Bivariate Plot of Grade 12 by Grade 11 T-Test Statistics for Mean Difference of STD-AFQT Score: Form 19G versus Reference Form: Sample #2..... | S-47 |

Table 1

**ASVAB Tests, Numbers of Items, Time Limits,
Normative Means, and Standard Deviations**

| <u>Test</u> | <u>No. Items</u> | <u>Time (min.)</u> | <u>Mean</u> | <u>S.D.</u> |
|--------------------------------|-----------------------------|-------------------------------|--------------------|--------------------|
| General Science (GS) | 25 | 11 | 15.950 | 5.010 |
| Arithmetic Reasoning (AR) | 30 | 36 | 18.009 | 7.373 |
| Word Knowledge (WK) | 35 | 11 | 26.270 | 7.710 |
| Paragraph Comprehension (PC) | 15 | 13 | 11.011 | 3.355 |
| Numerical Operations (NO) | 50 | 3 | 37.236 | 10.800 |
| Coding Speed (CS) | 84 | 7 | 47.606 | 16.763 |
| Auto and Shop Information (AS) | 25 | 11 | 14.317 | 5.550 |
| Mathematics Knowledge (MK) | 25 | 24 | 13.578 | 6.393 |
| Mechanical Comprehension (MC) | 25 | 19 | 14.165 | 5.349 |
| Electronics Information (EI) | 20 | 9 | 11.569 | 4.236 |
| Verbal (VE = WK + PC) | 50 | - | 37.281 | 10.595 |

Table 2
Sample 1:
Number and Percentages of Subjects by Test Form and Edit Procedure

| ASVAB Form | <u>Edit Stage</u> | | | |
|------------------------------|--------------------------|---|-----------------------------------|----------------------------------|
| | After Data File Matching | After Session Deletes (Balanced Data Set) | After Deletes for Invalid Version | After Deletes for "Below Chance" |
| REF | 59444 19.13% | 43684 19.92% | 43684 19.92% | 40566 19.77% |
| 18F | 65010 20.92% | 45034 20.54% | 45034 20.54% | 42012 20.47% |
| 18G | 64063 20.61% | 44398 20.25% | 44398 20.25% | 41618 20.28% |
| 19F | 62172 20.00% | 43332 19.76% | 43332 19.76% | 40576 19.77% |
| 19G | 60112 19.34% | 42837 19.53% | 42837 19.53% | 40464 19.72% |
| Missing | 8 0.00% | 6 0.00% | 0 | 0 |
| Total | 310809 | 219291 | 219285 | 205236 |
| Number Deleted at Edit Stage | ... | 91518 | 6 | 14049 |
| % of 310809 Deleted | | 29.45% | 0.00% | 4.52% |

Table 3
Sample 2:
Number and Percentages of Subjects by Test Form and Edit Procedure

| ASVAB Form | <u>Edit Stage</u> | | | |
|------------------------------|-------------------------|---|-----------------------------------|----------------------------------|
| | After Data File Merging | After Session Deletes (Balanced Data Set) | After Deletes for Invalid Version | After Deletes for "Below Chance" |
| REF | 161436 17.72% | 110576 19.91% | 110576 19.91% | 102694 19.78% |
| 18F | 193463 21.23% | 113862 20.50% | 113862 20.50% | 106168 20.44% |
| 18G | 190795 20.94% | 112489 20.26% | 112489 20.26% | 105467 20.31% |
| 19F | 185966 20.41% | 109814 19.78% | 109814 19.78% | 102672 19.77% |
| 19G | 179632 19.71% | 108561 19.55% | 108561 19.55% | 102300 19.70% |
| Missing | 51 0.00% | 49 0.00% | 0 | 0 |
| Total | 911343 | 555351 | 555302 | 519301 |
| Number Deleted at Edit Stage | ... | 355992 | 49 | 36001 |
| % of 911343 Deleted | | 39.06% | 0.00% | 3.95% |

Table 4a
Sample 1:
Number and Percentages of Subjects by Gender

| <u>Gender</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|---------------|-------------------|------------|------------|------------|------------|--------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| Female | | | | | | |
| Number | 21099 | 21738 | 21470 | 21007 | 20593 | 105907 |
| Percent | 48.30 | 48.27 | 48.36 | 48.48 | 48.08 | 48.3 |
| Male | | | | | | |
| Number | 22585 | 23296 | 22928 | 22325 | 22242 | 113376 |
| Percent | 51.70 | 51.73 | 51.64 | 51.52 | 51.92 | 51.7 |
| Total | | | | | | |
| Number | 43684 | 45034 | 44398 | 43332 | 42835 | 219283 |
| Percent | 19.92 | 20.54 | 20.25 | 19.76 | 19.53 | 100 |

Gender X Form: Chi-Square = 1.499 (d.f. = 4, pr. = 0.827)
 Effective Sample Size = 219283 Frequency Missing = 8

Table 4b
Sample 1:
Number and Percentages of Subjects by Education

| <u>Education Level</u> | <u>REF</u> | <u>ASVAB Form</u> | | | | <u>Total</u> |
|------------------------|------------|-------------------|------------|------------|------------|--------------|
| | | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| Grade 9 | | | | | | |
| Number | 2 | 2 | 2 | 1 | 2 | 9 |
| Percent | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Grade 10 | | | | | | |
| Number | 5673 | 5845 | 5805 | 5627 | 5720 | 28670 |
| Percent | 12.99 | 12.98 | 13.08 | 12.99 | 13.35 | 13.07 |
| Grade 11 | | | | | | |
| Number | 25537 | 26051 | 25844 | 25330 | 24802 | 127564 |
| Percent | 58.46 | 57.85 | 58.21 | 58.46 | 57.9 | 58.17 |
| Grade 12 | | | | | | |
| Number | 12383 | 13056 | 12663 | 12296 | 12244 | 62642 |
| Percent | 28.35 | 28.99 | 28.52 | 28.38 | 28.58 | 28.57 |
| Grade 13 | | | | | | |
| Number | 51 | 36 | 28 | 28 | 36 | 179 |
| Percent | 0.12 | 0.08 | 0.06 | 0.06 | 0.08 | 0.08 |
| Grade 14 | | | | | | |
| Number | 15 | 20 | 27 | 25 | 19 | 106 |
| Percent | 0.03 | 0.04 | 0.06 | 0.06 | 0.04 | 0.05 |
| Grade 15+ | | | | | | |
| Number | 23 | 24 | 26 | 24 | 13 | 110 |
| Percent | 0.05 | 0.05 | 0.06 | 0.06 | 0.03 | 0.05 |
| Total | | | | | | |
| Number | 43684 | 45034 | 44395 | 43331 | 42836 | 219280 |
| Percent | 19.92 | 20.54 | 20.25 | 19.76 | 19.53 | 100 |

Education Level X Form: Chi-Square = 29.245 (d.f. = 24, pr. = 0.211)

Effective Sample Size = 219280 Frequency Missing = 10

Table 4c
Sample 1:
Number and Percentages of Subjects by Race

| <u>Race</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|--------------|-------------------|------------|------------|------------|------------|--------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| Amer. Indian | | | | | | |
| Number | 929 | 994 | 974 | 976 | 939 | 4812 |
| Percent | 2.13 | 2.21 | 2.19 | 2.25 | 2.19 | 2.19 |
| Hispanic | | | | | | |
| Number | 2767 | 2870 | 2886 | 2787 | 2716 | 14026 |
| Percent | 6.33 | 6.37 | 6.5 | 6.43 | 6.34 | 6.4 |
| Asian | | | | | | |
| Number | 566 | 623 | 595 | 574 | 560 | 2918 |
| Percent | 1.3 | 1.38 | 1.34 | 1.32 | 1.31 | 1.33 |
| Black | | | | | | |
| Number | 5850 | 6033 | 5980 | 5903 | 5792 | 29558 |
| Percent | 13.39 | 13.4 | 13.47 | 13.62 | 13.52 | 13.48 |
| White | | | | | | |
| Number | 32966 | 33944 | 33325 | 32443 | 32258 | 164936 |
| Percent | 75.47 | 75.38 | 75.07 | 74.88 | 75.31 | 75.22 |
| Other | | | | | | |
| Number | 601 | 567 | 632 | 645 | 569 | 3014 |
| Percent | 1.38 | 1.26 | 1.42 | 1.49 | 1.33 | 1.37 |
| Total | | | | | | |
| Number | 43679 | 45031 | 44392 | 43328 | 42834 | 219264 |
| Percent | 19.92 | 20.54 | 20.25 | 19.76 | 19.54 | 100 |

Race X Form: Chi-Square = 17.010 (d.f. = 20. pr. = 0.652)
 Effective Sample Size = 219264 Frequency Missing = 27

Table 5a
Sample 2:
Number and Percentages of Subjects by Gender

| <u>Gender</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|---------------|-------------------|------------|------------|------------|------------|--------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| Female | | | | | | |
| Number | 53821 | 55094 | 54174 | 53290 | 52289 | 268668 |
| Percent | 48.67 | 48.39 | 48.16 | 48.53 | 48.17 | 48.38 |
| Male | | | | | | |
| Number | 56752 | 58765 | 58314 | 56524 | 56270 | 286625 |
| Percent | 51.33 | 51.61 | 51.84 | 51.47 | 51.83 | 51.62 |
| Total | | | | | | |
| Number | 110573 | 113859 | 112488 | 109814 | 108559 | 555293 |
| Percent | 19.91 | 20.50 | 20.26 | 19.78 | 19.55 | 100.00 |

Gender X Form: Chi-Square = 8.968 (d.f. = 4, pr. = 0.062)
 Effective Sample Size = 555293 Frequency Missing = 58

Table 5b

**Sample 2:
Number and Percentages of Subjects by Education**

| <u>Education Level</u> | <u>ASVAB Form</u> | | | | | |
|------------------------|-------------------|------------|------------|------------|------------|--------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | <u>Total</u> |
| Grade 9 | | | | | | |
| Number | 15 | 17 | 13 | 14 | 18 | 77 |
| Percent | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 |
| Grade 10 | | | | | | |
| Number | 12794 | 13217 | 13090 | 12713 | 12899 | 64713 |
| Percent | 11.57 | 11.61 | 11.64 | 11.58 | 11.88 | 11.65 |
| Grade 11 | | | | | | |
| Number | 64723 | 66133 | 65304 | 64065 | 63089 | 323314 |
| Percent | 58.53 | 58.08 | 58.05 | 58.34 | 58.11 | 58.22 |
| Grade 12 | | | | | | |
| Number | 32855 | 34308 | 33901 | 32849 | 32400 | 166313 |
| Percent | 29.71 | 30.13 | 30.14 | 29.91 | 29.85 | 29.95 |
| Grade 13 | | | | | | |
| Number | 107 | 80 | 68 | 77 | 86 | 418 |
| Percent | 0.10 | 0.07 | 0.06 | 0.07 | 0.08 | 0.08 |
| Grade 14 | | | | | | |
| Number | 33 | 46 | 54 | 35 | 35 | 203 |
| Percent | 0.03 | 0.04 | 0.05 | 0.03 | 0.03 | 0.04 |
| Grade 15+ | | | | | | |
| Number | 47 | 60 | 57 | 60 | 33 | 257 |
| Percent | 0.04 | 0.05 | 0.05 | 0.05 | 0.03 | 0.05 |
| Total | | | | | | |
| Number | 110574 | 113861 | 112487 | 109813 | 108560 | 555295 |
| Percent | 19.91 | 20.50 | 20.26 | 19.78 | 19.55 | 100 |

Education Level X Form: Chi-Square = 43.368 (d.f. = 24, pr. = 0.009)

Effective Sample Size = 555295 Frequency Missing = 56

Table 5c
Sample 2:
Number and Percentages of Subjects by Race

| <u>Race</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|---------------------|-------------------|------------|------------|------------|------------|--------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| Amer. Indian | | | | | | |
| Number | 2067 | 2211 | 2125 | 2095 | 2103 | 10601 |
| Percent | 1.87 | 1.94 | 1.89 | 1.91 | 1.94 | 1.91 |
| Hispanic | | | | | | |
| Number | 7511 | 7828 | 7760 | 7481 | 7361 | 37941 |
| Percent | 6.79 | 6.87 | 6.90 | 6.81 | 6.78 | 6.83 |
| Asian | | | | | | |
| Number | 1586 | 1739 | 1647 | 1649 | 1595 | 8216 |
| Percent | 1.43 | 1.53 | 1.46 | 1.50 | 1.47 | 1.48 |
| Black | | | | | | |
| Number | 16184 | 16566 | 16541 | 16282 | 16075 | 81648 |
| Percent | 14.64 | 14.55 | 14.70 | 14.83 | 14.81 | 14.70 |
| White | | | | | | |
| Number | 81506 | 83808 | 82679 | 80524 | 79787 | 408304 |
| Percent | 73.71 | 73.60 | 73.50 | 73.33 | 73.50 | 73.53 |
| Other | | | | | | |
| Number | 1712 | 1692 | 1721 | 1767 | 1634 | 8526 |
| Percent | 1.55 | 1.49 | 1.53 | 1.61 | 1.51 | 1.54 |
| Not Appl | | | | | | |
| Number | 0 | 2 | 0 | 2 | 0 | 4 |
| Percent | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Unknown | | | | | | |
| Number | 10 | 16 | 16 | 14 | 6 | 62 |
| Percent | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| Total | | | | | | |
| Number | 43679 | 45031 | 44392 | 43328 | 42834 | 219264 |
| Percent | 19.92 | 20.54 | 20.25 | 19.76 | 19.54 | 100 |

Race X Form: Chi-Square = 31.243 (d.f. = 28 pr. = 0.306)
 Effective Sample Size = 555302 Frequency Missing = 49

Table 6a
Sample 1: Grade 12
Number and Percentages of Subjects by Gender

| <u>Gender</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|---------------|-------------------|----------------|----------------|----------------|----------------|----------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| Female | 5342 46.59 | 5771 47.43 | 5593 47.14 | 5426 47.12 | 5496 47.46 | 27628 47.15 |
| Male | 61255 53.41 | 6397 52.57 | 6272 52.86 | 6090 52.88 | 6085 52.54 | 30969 52.85 |
| Total | 11467 19.57 | 12168 20.77 | 11865 20.25 | 11516 19.65 | 11581 19.76 | 58597 100 |

Gender X Form: Chi-Square = 2.285 (d.f. = 4, pr. = 0.684)
 Effective Sample Size = 58597 Frequency Missing = 0

Table 6b
Sample 1: Grade 12
Number and Percentages of Subjects
by Population Group

| <u>Population Group</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|-------------------------|-------------------|----------------|----------------|----------------|----------------|-----------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| American Indian | 274 2.39 | 322 2.65 | 289 2.44 | 267 2.32 | 288 2.49 | 1440 2.46 |
| Hispanic | 861 7.51 | 935 7.68 | 907 7.64 | 917 7.96 | 893 7.71 | 4513 7.70 |
| Asian | 181 1.58 | 212 1.74 | 190 1.60 | 199 1.73 | 190 1.64 | 972 1.66 |
| Black | 2001 17.45 | 2233 18.35 | 2246 18.93 | 2171 18.85 | 2173 18.76 | 10824 18.47 |
| White | 7996 69.73 | 8317 68.35 | 8058 67.91 | 7767 67.45 | 7882 68.06 | 40020 68.30 |
| Other | 154 1.34 | 149 1.22 | 175 1.47 | 195 1.69 | 155 1.34 | 828 1.41 |
| Total | 11467 19.57 | 12168 20.77 | 11865 20.25 | 11516 19.65 | 11581 19.76 | 58597 100.00 |

Statistic DF Value Prob
Chi-Square 20 31.155 0.053
Sample Size = 58597

Table 6c

**Sample 1: Grade 12
Number and Percentages of Subjects
by Post-High School Intention**

| <u>Post-High School Intention</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|---------------------------------------|-------------------|----------------|----------------|----------------|----------------|-----------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| 4yr.Col | 5365 46.83 | 5619 46.24 | 5540 46.72 | 5521 47.98 | 5448 47.08 | 27493 46.96 |
| 2yr.Col | 1279 11.16 | 1300 10.70 | 1238 10.44 | 1231 10.70 | 1209 10.45 | 6257 10.69 |
| VoTech | 555 4.84 | 575 4.73 | 602 5.08 | 509 4.42 | 561 4.85 | 2802 4.79 |
| Military | 1361 11.88 | 1448 11.92 | 1339 11.29 | 1338 11.63 | 1328 11.48 | 6814 11.64 |
| Work | 401 3.50 | 473 3.89 | 458 3.86 | 400 3.48 | 454 3.92 | 2186 3.73 |
| Undecided | 2496 21.79 | 2736 22.52 | 2682 22.62 | 2508 21.80 | 2571 22.22 | 12993 22.19 |
| Total | 11457 19.57 | 12151 20.75 | 11859 20.26 | 11507 19.65 | 11571 19.76 | 58545 100.00 |

Statistic DF Value Prob
 Chi-Square 20 25.599 0.179

Effective Sample Size = 58545

Frequency Missing = 52

Table 7a

Sample 1: Grade 11
Number and Percentages of Subjects by Gender

| <u>Gender</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|---------------|-------------------|----------------|----------------|----------------|----------------|------------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| Female | 11735 49.08 | 11927 48.70 | 11966 49.07 | 11725 49.13 | 11509 48.84 | 58862 48.96 |
| Male | 12176 50.92 | 12562 51.30 | 12420 50.93 | 12142 50.87 | 12055 51.16 | 61355 51.04 |
| Total | 23911 19.89 | 24489 20.37 | 24386 20.28 | 23867 19.85 | 23564 19.60 | 120217 100.00 |

Statistic DF Value Prob
Chi-Square 4 1.290 0.863

Effective Sample Size = 120217

Frequency Missing = 1

Table 7b

**Sample 1: Grade 11
Number and Percentages of Subjects
by Population Group**

| <u>Population Group</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|-------------------------|-------------------|----------------|----------------|----------------|----------------|------------------|
| | <u>18E</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| American Indian | 473 1.98 | 488 1.99 | 504 2.07 | 512 2.15 | 474 2.01 | 2451 2.04 |
| Hispanic | 1195 5.00 | 1217 4.97 | 1304 5.35 | 1221 5.12 | 1233 5.23 | 6170 5.13 |
| Asian | 264 1.10 | 295 1.20 | 289 1.19 | 270 1.13 | 281 1.19 | 1399 1.16 |
| Black | 2356 9.85 | 2467 10.07 | 2477 10.16 | 2478 10.38 | 2432 10.32 | 12210 10.16 |
| White | 19310 80.76 | 19738 80.60 | 19505 79.98 | 19083 79.96 | 18859 80.03 | 96495 80.27 |
| Other | 313 1.31 | 284 1.16 | 307 1.26 | 303 1.27 | 286 1.21 | 1493 1.24 |
| Total | 23911 19.89 | 24489 20.37 | 24386 20.28 | 23867 19.85 | 23565 19.60 | 120218 100.00 |

| | | | |
|----------------------|----|--------|-------|
| Statistic | DF | Value | Prob |
| Chi-Square | 20 | 16.989 | 0.654 |
| Sample Size = 120218 | | | |

Table 7c

Sample 1: Grade 11 Number and Percentages of Subjects by Post-High School Intention

| <u>Post-High School Intention</u> | <u>ASVAB Form</u> | | | | | |
|-----------------------------------|-------------------|----------------|----------------|----------------|----------------|------------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | <u>Total</u> |
| 4yr.Col | 12595 52.72 | 12813 52.36 | 12649 51.90 | 12434 52.13 | 12294 52.21 | 62785 52.26 |
| 2yr.Col | 1606 6.72 | 1674 6.84 | 1679 6.89 | 1696 7.11 | 1592 6.76 | 8247 6.86 |
| VoTech | 973 4.07 | 925 3.78 | 990 4.06 | 920 3.86 | 954 4.05 | 4762 3.96 |
| Military | 2126 8.90 | 2231 9.12 | 2162 8.87 | 2131 8.93 | 2147 9.12 | 10797 8.99 |
| Work | 694 2.90 | 709 2.90 | 674 2.77 | 724 3.04 | 692 2.94 | 3493 2.91 |
| Undecided | 5897 24.68 | 6121 25.01 | 6220 25.52 | 5949 24.94 | 5869 24.92 | 30056 25.02 |
| Total | 23891 19.89 | 24473 20.37 | 24374 20.29 | 23854 19.86 | 23548 19.60 | 120140 100.00 |

| Statistic | DF | Value | Prob |
|------------|----|--------|-------|
| Chi-Square | 20 | 17.806 | 0.600 |

Effective Sample Size = 120140

Frequency Missing = 78

Table 8a
Sample 2: Grade 12
Number and Percentages of Subjects by Gender

| <u>Gender</u> | <u>ASVAB Form</u> | | | | | |
|---------------|-------------------|----------------|----------------|----------------|----------------|------------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | <u>Total</u> |
| Female | 14345 47.18 | 15211 47.49 | 14895 46.91 | 14564 47.39 | 14487 47.35 | 73502 47.26 |
| Male | 16057 52.82 | 16822 52.51 | 16859 53.09 | 16167 52.61 | 16108 52.65 | 82013 52.74 |
| Total | 30402 19.55 | 32033 20.60 | 31754 20.42 | 30731 19.76 | 30595 19.67 | 155515 100.00 |

Statistic DF Value Prob
 Chi-Square 4 2.621 0.623
 Sample Size = 155515

Table 8b
Sample 2: Grade 12
Number and Percentages of Subjects
by Population Group

| <u>Population Group</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|-------------------------|-------------------|----------------|----------------|----------------|----------------|------------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| American Indian | 500 1.64 | 594 1.85 | 567 1.79 | 507 1.65 | 569 1.86 | 2737 1.76 |
| Hispanic | 2554 8.40 | 2743 8.56 | 2653 8.35 | 2630 8.56 | 2604 8.51 | 13184 8.48 |
| Asian | 503 1.65 | 556 1.74 | 535 1.68 | 537 1.75 | 497 1.62 | 2628 1.69 |
| Black | 5739 18.88 | 6216 19.40 | 6365 20.04 | 6194 20.16 | 6165 20.15 | 30679 19.73 |
| White | 20617 67.81 | 21392 66.78 | 21102 66.45 | 20259 65.92 | 20247 66.18 | 103617 66.63 |
| Other | 489 1.61 | 532 1.66 | 532 1.68 | 604 1.97 | 513 1.68 | 2670 1.72 |
| Total | 30402 19.55 | 32033 20.60 | 31754 20.42 | 30731 19.76 | 30595 19.67 | 155515 100.00 |

Statistic
Chi-Square DF Value
 20 55.539 Prob
 0.000
Sample Size = 155515

Table 8c

**Sample 2: Grade 12
Number and Percentages of Subjects
by Post-High School Intention**

**Post-High School
Intention**

ASVAB Form

| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | <u>Total</u> |
|-----------|-------------------|-------------------|-------------------|-------------------|-------------------|---------------------|
| 4yr.Col | 14153 46.55 | 14741 46.02 | 14674 46.21 | 14612 47.55 | 14392 47.04 | 72572 46.67 |
| 2yr.Col | 3322 10.93 | 3486 10.88 | 3355 10.57 | 3225 10.49 | 3126 10.22 | 16514 10.62 |
| VoTech | 1435 4.72 | 1536 4.80 | 1516 4.77 | 1390 4.52 | 1493 4.88 | 7370 4.74 |
| Military | 3776 12.42 | 3996 12.47 | 3796 11.95 | 3762 12.24 | 3729 12.19 | 19059 12.26 |
| Work | 985 3.24 | 1101 3.44 | 1076 3.39 | 982 3.20 | 1072 3.50 | 5216 3.35 |
| Undecided | 6731 22.14 | 7173 22.39 | 7337 23.11 | 6760 22.00 | 6783 22.17 | 34645 22.27 |
| Total | 30402 19.55 | 32033 20.60 | 31754 20.42 | 30731 19.76 | 30595 19.67 | 155515 100.00 |

Statistic

DF

Value

Prob

Chi-Square

20

46.589

0.001

Effective Sample Size = 155462

Frequency Missing = 53

Table 9a
Sample 2: Grade 11
Number and Percentages of Subjects by Gender

| <u>Gender</u> | <u>ASVAB Form</u> | | | | | |
|---------------|-------------------|------------|------------|------------|------------|--------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | <u>Total</u> |
| Female | 29886 | 30309 | 30160 | 29689 | 29196 | 149240 |
| | 49.30 | 48.82 | 48.91 | 49.22 | 48.79 | 49.01 |
| Male | 30730 | 31775 | 31506 | 30634 | 30640 | 155285 |
| | 50.70 | 51.18 | 51.09 | 50.78 | 51.21 | 50.99 |
| Total | 60616 | 62084 | 61666 | 60323 | 59836 | 304525 |
| | 19.91 | 20.39 | 20.25 | 19.81 | 19.65 | 100.00 |

| Statistic | DF | Value | Prob |
|------------|----|-------|-------|
| Chi-Square | 4 | 5.405 | 0.248 |

Chi-Square
Effective Sample Size = 304525

Effective Sample Size =
Frequency Missing = ?

Table 9b

**Sample 2: Grade 11
Number and Percentages of Subjects
by Population Group**

| <u>Population Group</u> | <u>ASVAB Form</u> | | | | | <u>Total</u> |
|-------------------------|-------------------|----------------|----------------|----------------|----------------|------------------|
| | <u>18E</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | |
| American Indian | 1186 1.96 | 1205 1.94 | 1161 1.88 | 1149 1.90 | 1155 1.93 | 5856 1.92 |
| Hispanic | 3260 5.38 | 3354 5.40 | 3427 5.56 | 3200 5.30 | 3271 5.47 | 16512 5.42 |
| Asian | 767 1.27 | 871 1.40 | 788 1.28 | 786 1.30 | 844 1.41 | 4056 1.33 |
| Black | 6666 11.00 | 6990 11.26 | 7002 11.35 | 7000 11.60 | 6893 11.52 | 34551 11.35 |
| White | 47889 79.00 | 48846 78.68 | 48490 78.63 | 47407 78.59 | 46922 78.42 | 239554 78.66 |
| Other | 848 1.40 | 819 1.32 | 798 1.29 | 781 1.29 | 752 1.26 | 3998 1.31 |
| Total | 60616 19.90 | 62085 20.39 | 61666 20.25 | 60323 19.81 | 59837 19.65 | 304527 100.00 |

Statistic
Chi-Square
Sample Size = 304527

| | | |
|----|--------|-------|
| DF | Value | Prob |
| 20 | 32.727 | 0.036 |

Table 9c

Sample 2: Grade 11 Number and Percentages of Subjects by Post-High School Intention

| <u>Post-High School Intention</u> | <u>ASVAB Form</u> | | | | | |
|---------------------------------------|-------------------|----------------|----------------|----------------|----------------|------------------|
| | <u>REF</u> | <u>18F</u> | <u>18G</u> | <u>19F</u> | <u>19G</u> | <u>Total</u> |
| 4yr.Col | 32007 52.80 | 32699 52.67 | 32226 52.26 | 31761 52.65 | 31459 52.57 | 160152 52.59 |
| 2yr.Col | 4081 6.73 | 4238 6.83 | 4057 6.58 | 4243 7.03 | 4036 6.74 | 20655 6.78 |
| VoTech | 2430 4.01 | 2364 3.81 | 2450 3.97 | 2380 3.95 | 2429 4.06 | 12053 3.96 |
| Military | 5772 9.52 | 5922 9.54 | 5827 9.45 | 5598 9.28 | 5598 9.36 | 28717 9.43 |
| Work | 1760 2.90 | 1750 2.82 | 1757 2.85 | 1708 2.83 | 1771 2.96 | 8746 2.87 |
| Undecided | 14534 24.03 | 15074 24.34 | 15315 24.89 | 14606 24.26 | 14510 24.31 | 74039 24.37 |
| Total | 60616 19.90 | 62085 20.39 | 61666 20.25 | 60323 19.81 | 59837 19.65 | 304527 100.00 |

| Statistic | DF | Value | Prob |
|------------|----|--------|-------|
| Chi-Square | 20 | 33.570 | 0.029 |

Effective Sample Size = 304452

Frequency Missing = 75

Table 10

Sample 1: Grade 12
Test Raw Score Mean, Standard Deviation, Skewness, Kurtosis,
and Number of Subjects by Form

| GS | | REF | 18F | 18G | 19F | 19G |
|------|-----------|------------|------------|------------|------------|------------|
| Form | Items | | | | | |
| N | 11467 | | 12166 | 11865 | 11516 | 11580 |
| Mean | 15.311851 | | 16.646638 | 16.482764 | 16.602900 | 16.454318 |
| SD | 4.085538 | | 4.237231 | 4.357388 | 3.734747 | 3.808304 |
| Skew | -0.092139 | | -0.209590 | -0.169507 | -0.246118 | -0.227674 |
| Kurt | 2.466295 | | 2.389749 | 2.304373 | 2.613092 | 2.575612 |
| Deg | 4 | | 7 | 6 | 4 | 10 |
| AR | | REF | 18F | 18G | 19F | 19G |
| Form | Items | | | | | |
| N | 11467 | | 12168 | 11865 | 11516 | 11581 |
| Mean | 17.560914 | | 18.698060 | 18.150527 | 18.784213 | 18.173646 |
| SD | 6.026599 | | 5.804626 | 5.748333 | 5.873114 | 5.752365 |
| Skew | 0.177629 | | 0.071930 | 0.056591 | 0.045507 | 0.044117 |
| Kurt | 2.199775 | | 2.171925 | 2.296487 | 2.165450 | 2.282319 |
| Deg | 6 | | 6 | 7 | 10 | 10 |
| WK | | REF | 18F | 18G | 19F | 19G |
| Form | Items | | | | | |
| N | 11467 | | 12167 | 11865 | 11516 | 11580 |
| Mean | 25.623877 | | 25.919947 | 25.801854 | 26.066777 | 25.792660 |
| SD | 5.609582 | | 5.961508 | 5.521816 | 5.996842 | 5.563351 |
| Skew | -0.586849 | | -0.616481 | -0.631167 | -0.646639 | -0.616236 |
| Kurt | 2.877346 | | 2.883465 | 3.102073 | 2.883024 | 2.993662 |
| Deg | 6 | | 7 | 6 | 10 | 10 |
| PC | | REF | 18F | 18G | 19F | 19G |
| Form | Items | | | | | |
| N | 11463 | | 12161 | 11863 | 11510 | 11576 |
| Mean | 10.942075 | | 10.883398 | 11.109079 | 10.360209 | 10.809779 |
| SD | 2.903874 | | 3.243822 | 3.048553 | 3.370395 | 3.127380 |
| Skew | -0.813296 | | -0.723241 | -0.735142 | -0.499022 | -0.640999 |
| Kurt | 3.135402 | | 2.681398 | 2.853575 | 2.325397 | 2.664515 |
| Deg | 4 | | 6 | 5 | 8 | 8 |

Table 10
continued

NO

Statistics:

| | REF | 18F | 18G | 19F | 19G |
|-------|------------|------------|------------|------------|------------|
| Form | | | | | |
| Items | 50 | 50 | 50 | 50 | 50 |
| N | 11465 | 12165 | 11863 | 11515 | 11577 |
| Mean | 38.544701 | 37.106453 | 36.680182 | 37.648893 | 37.452276 |
| SD | 8.572379 | 8.712295 | 8.768312 | 8.573108 | 8.734328 |
| Skew | -0.461387 | -0.339964 | -0.283163 | -0.383017 | -0.385880 |
| Kurt | 2.604572 | 2.428117 | 2.340243 | 2.514189 | 2.479902 |
| Deg | 10 | 10 | 10 | 10 | 10 |

CS

Statistics:

| | REF | 18F | 18G | 19F | 19G |
|-------|------------|------------|------------|------------|------------|
| Form | | | | | |
| Items | 84 | 84 | 84 | 84 | 84 |
| N | 11467 | 12167 | 11863 | 11516 | 11580 |
| Mean | 51.685358 | 50.914523 | 51.138245 | 51.529785 | 51.545337 |
| SD | 12.971660 | 13.036422 | 13.097168 | 13.315698 | 13.316770 |
| Skew | -0.067972 | -0.009900 | -0.122754 | -0.007441 | -0.014916 |
| Kurt | 3.292145 | 3.237612 | 3.297829 | 3.103338 | 3.202458 |
| Deg | 8 | 9 | 8 | 9 | 8 |

AS

Statistics:

| | REF | 18F | 18G | 19F | 19G |
|-------|------------|------------|------------|------------|------------|
| Form | | | | | |
| Items | 25 | 25 | 25 | 25 | 25 |
| N | 11467 | 12162 | 11857 | 11515 | 11581 |
| Mean | 12.185750 | 11.574741 | 11.622839 | 12.468519 | 12.325533 |
| SD | 4.431866 | 4.769653 | 4.729357 | 4.529743 | 4.510173 |
| Skew | 0.514744 | 0.558377 | 0.562379 | 0.662364 | 0.626285 |
| Kurt | 2.733004 | 2.694043 | 2.743043 | 2.820889 | 2.810514 |
| Deg | 9 | 8 | 10 | 5 | 9 |

MK

Statistics:

| | REF | 18F | 18G | 19F | 19G |
|-------|------------|------------|------------|------------|------------|
| Form | | | | | |
| Items | 25 | 25 | 25 | 25 | 25 |
| N | 11466 | 12165 | 11864 | 11513 | 11581 |
| Mean | 15.136054 | 15.004439 | 14.948837 | 15.101885 | 15.002072 |
| SD | 5.520143 | 5.691463 | 5.457862 | 5.719063 | 5.488423 |
| Skew | -0.026623 | -0.035743 | 0.064315 | -0.072322 | 0.061620 |
| Kurt | 1.980440 | 1.996001 | 2.013123 | 1.992550 | 1.989285 |
| Deg | 7 | 7 | 8 | 10 | 7 |

Table 10
continued

**MC
Statistics:**

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 25 | 25 | 25 | 25 | 25 |
| N | 11456 | 12150 | 11852 | 11504 | 11559 |
| Mean | 12.926501 | 13.690453 | 13.512234 | 13.409944 | 13.063846 |
| SD | 4.746821 | 4.704160 | 4.721569 | 4.919879 | 4.857364 |
| Skew | 0.339750 | 0.092173 | 0.135702 | 0.226883 | 0.246000 |
| Kurt | 2.331825 | 2.305192 | 2.320582 | 2.234753 | 2.296017 |
| Deg | 8 | 9 | 4 | 7 | 8 |

EI Statistics:

| Form | REF | 18F | 18G | 19F | 19G |
|-------|----------|-----------|-----------|----------|----------|
| Items | 20 | 20 | 20 | 20 | 20 |
| N | 11438 | 12130 | 11839 | 11478 | 11544 |
| Mean | 9.826368 | 10.350124 | 10.252809 | 9.762415 | 9.639466 |
| SD | 3.485453 | 3.338575 | 3.274299 | 3.407927 | 3.321316 |
| Skew | 0.198581 | 0.301239 | 0.330147 | 0.327882 | 0.382746 |
| Kurt | 2.503084 | 2.744388 | 2.794506 | 2.823345 | 2.930013 |
| Deg | 4 | 9 | 7 | 4 | 10 |

**VE
Statistics:**

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 50 | 50 | 50 | 50 | 50 |
| N | 11467 | 12168 | 11865 | 11516 | 11581 |
| Mean | 36.562135 | 36.794954 | 36.909060 | 36.421587 | 36.595544 |
| SD | 7.770092 | 8.512258 | 7.917734 | 8.690723 | 8.032291 |
| Skew | -0.615896 | -0.596899 | -0.600797 | -0.528464 | -0.560765 |
| Kurt | 2.860546 | 2.742702 | 2.868390 | 2.575149 | 2.740677 |
| Deg | 7 | 10 | 6 | 10 | 10 |

Table 11

Sample 1: Grade 11
Test Raw Score Mean, Standard Deviation, Skewness, Kurtosis,
and Number of Subjects by Form

GS Statistics:

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 25 | 25 | 25 | 25 | 25 |
| N | 23910 | 24488 | 24386 | 23865 | 23563 |
| Mean | 15.611920 | 17.087471 | 17.018617 | 16.977331 | 16.983449 |
| SD | 3.944958 | 4.097366 | 4.141987 | 3.648882 | 3.674911 |
| Skew | -0.115953 | -0.289773 | -0.262045 | -0.333708 | -0.345081 |
| Kurt | 2.499757 | 2.486019 | 2.419410 | 2.698696 | 2.733494 |
| Deg | 10 | 6 | 10 | 7 | 7 |

AR Statistics:

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 30 | 30 | 30 | 30 | 30 |
| N | 23911 | 24489 | 24386 | 23867 | 23565 |
| Mean | 17.765004 | 18.853077 | 18.507873 | 18.750450 | 18.510460 |
| SD | 6.009022 | 5.766013 | 5.700251 | 5.851587 | 5.655041 |
| Skew | 0.131599 | 0.030790 | 0.002331 | 0.049207 | -0.016211 |
| Kurt | 2.141514 | 2.180123 | 2.292048 | 2.145033 | 2.301435 |
| Deg | 10 | 7 | 7 | 10 | 10 |

WK Statistics:

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 35 | 35 | 35 | 35 | 35 |
| N | 23911 | 24487 | 24386 | 23867 | 23565 |
| Mean | 25.632429 | 25.969862 | 25.894858 | 26.096032 | 25.935243 |
| SD | 5.424656 | 5.731249 | 5.296082 | 5.773739 | 5.277390 |
| Skew | -0.589781 | -0.627762 | -0.640203 | -0.653528 | -0.685909 |
| Kurt | 2.970689 | 2.988523 | 3.163814 | 2.987787 | 3.308439 |
| Deg | 6 | 10 | 10 | 7 | 10 |

PC Statistics:

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 15 | 15 | 15 | 15 | 15 |
| N | 23906 | 24484 | 24383 | 23848 | 23554 |
| Mean | 11.018113 | 10.934079 | 11.216421 | 10.408336 | 10.954912 |
| SD | 2.875910 | 3.233415 | 3.023422 | 3.367500 | 3.093077 |
| Skew | -0.828295 | -0.755615 | -0.803077 | -0.548463 | -0.703461 |
| Kurt | 3.148495 | 2.764174 | 3.001513 | 2.408701 | 2.795754 |
| Deg | 8 | 8 | 5 | 8 | 8 |

Table 11
continued

NO Statistics:

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 50 | 50 | 50 | 50 | 50 |
| N | 23908 | 24487 | 24382 | 23865 | 23564 |
| Mean | 38.100845 | 36.782538 | 36.293659 | 37.158056 | 37.008954 |
| SD | 8.413626 | 8.563445 | 8.624484 | 8.524528 | 8.606663 |
| Skew | -0.360810 | -0.251303 | -0.206034 | -0.288211 | -0.293810 |
| Kurt | 2.459603 | 2.363576 | 2.372187 | 2.415549 | 2.449423 |
| Deg | 10 | 10 | 10 | 10 | 10 |

CS Statistics:

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 84 | 84 | 84 | 84 | 84 |
| N | 23907 | 24488 | 24382 | 23866 | 23562 |
| Mean | 50.183712 | 49.274951 | 49.690632 | 50.012193 | 50.056235 |
| SD | 12.696645 | 12.693648 | 12.766087 | 12.861199 | 13.022716 |
| Skew | -0.076268 | -0.043653 | -0.107486 | 0.037902 | -0.016454 |
| Kurt | 3.364996 | 3.331152 | 3.335694 | 3.179372 | 3.237670 |
| Deg | 8 | 9 | 8 | 7 | 8 |

AS Statistics:

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 25 | 25 | 25 | 25 | 25 |
| N | 23907 | 24480 | 24376 | 23865 | 23564 |
| Mean | 11.789852 | 11.289747 | 11.401953 | 12.105007 | 12.105542 |
| SD | 4.222125 | 4.498310 | 4.512971 | 4.274535 | 4.280128 |
| Skew | 0.530666 | 0.581871 | 0.556845 | 0.631264 | 0.633282 |
| Kurt | 2.888298 | 2.834950 | 2.827673 | 2.977504 | 2.932955 |
| Deg | 5 | 8 | 10 | 7 | 9 |

MK Statistics:

| Form | REF | 18F | 18G | 19F | 19G |
|-------|-----------|-----------|-----------|-----------|-----------|
| Items | 25 | 25 | 25 | 25 | 25 |
| N | 23908 | 24483 | 24385 | 23865 | 23565 |
| Mean | 15.372511 | 15.345709 | 15.289891 | 15.332328 | 15.372841 |
| SD | 5.421109 | 5.580942 | 5.455018 | 5.602315 | 5.431284 |
| Skew | -0.083631 | -0.108504 | -0.035024 | -0.115214 | -0.029770 |
| Kurt | 2.040001 | 2.018312 | 2.007399 | 1.998252 | 1.994948 |
| Deg | 9 | 8 | 10 | 10 | 8 |

Table 11
continued

MC Statistics:

| | REF | 18F | 18G | 19F | 19G |
|-------|------------|------------|------------|------------|------------|
| Form | | | | | |
| Items | 25 | 25 | 25 | 25 | 25 |
| N | 23891 | 24462 | 24356 | 23833 | 23528 |
| Mean | 12.989954 | 13.778963 | 13.773362 | 13.439223 | 13.299600 |
| SD | 4.580961 | 4.577014 | 4.580555 | 4.779580 | 4.802780 |
| Skew | 0.309567 | 0.063493 | 0.059583 | 0.199880 | 0.190823 |
| Kurt | 2.371743 | 2.313454 | 2.328494 | 2.291603 | 2.262393 |
| Deg | 9 | 10 | 7 | 9 | 10 |

EI Statistics:

| | REF | 18F | 18G | 19F | 19G |
|-------|------------|------------|------------|------------|------------|
| Form | | | | | |
| Items | 20 | 20 | 20 | 20 | 20 |
| N | 23840 | 24412 | 24310 | 23779 | 23488 |
| Mean | 9.659606 | 10.178765 | 10.201563 | 9.510282 | 9.481267 |
| SD | 3.438242 | 3.220296 | 3.212278 | 3.234829 | 3.166326 |
| Skew | 0.184375 | 0.279324 | 0.280617 | 0.298557 | 0.349845 |
| Kurt | 2.528941 | 2.791250 | 2.783368 | 2.867716 | 2.968813 |
| Deg | 9 | 10 | 7 | 9 | 10 |

VE Statistics:

| | REF | 18F | 18G | 19F | 19G |
|-------|------------|------------|------------|------------|------------|
| Form | | | | | |
| Items | 50 | 50 | 50 | 50 | 50 |
| N | 23911 | 24489 | 24386 | 23867 | 23565 |
| Mean | 36.648237 | 36.899588 | 37.109899 | 36.496082 | 36.885041 |
| SD | 7.553958 | 8.297857 | 7.678266 | 8.477017 | 7.720998 |
| Skew | -0.630531 | -0.614029 | -0.649165 | -0.563452 | -0.634550 |
| Kurt | 2.942646 | 2.789119 | 2.977026 | 2.670417 | 2.958728 |
| Deg | 9 | 9 | 9 | 10 | 8 |

Table 12

Sample 1: Grade 12
Likelihood Ratio Chi-Square Tests of Significance Item Order Effects

| Subtest | Form | Degree of Polynomial | Chi-Square | D.F. | Probability |
|----------------|-------------|-----------------------------|-------------------|-------------|--------------------|
| GS | 18F&G | 7 (7 & 6) | 24.17904 | 7 | *4.00105965 |
| GS | 19F&G | 10 (4&10) | 21.67304 | 10 | 0.01685969 |
| AR | 18&19F | 10 (6&10) | 11.84217 | 10 | 0.29575865 |
| AR | 18&19G | 10 (7&10) | 9.16431 | 10 | 0.51658394 |
| WK | 18&19F | 10 (7&10) | 12.47867 | 10 | 0.25429654 |
| WK | 18&19G | 10 (6&10) | 7.54501 | 10 | 0.67318375 |
| PC | 18&19F | 8 (6 & 8) | 166.95303 | 8 | *0.00000000 |
| PC | 18&19G | 8 (5 & 8) | 58.02326 | 8 | *0.00000000 |
| CS | 18F&G | 9 (9 & 8) | 28.23786 | 9 | *0.00087036 |
| CS | 19F&G | 9 (9 & 8) | 8.59337 | 9 | 0.47562360 |
| AS | 18F&G | 10 (8&10) | 13.84327 | 10 | 0.18026141 |
| AS | 19F&G | 9 (5 & 9) | 26.36370 | 9 | 0.00178089 |
| MK | 18&19F | 10 (7&10) | 10.56614 | 10 | 0.39230274 |
| MK | 18&19G | 8 (8 & 7) | 5.10131 | 8 | 0.74669536 |
| MC | 18F&G | 9 (9 & 4) | 12.38885 | 9 | 0.19226493 |
| MC | 19F&G | 8 (7 & 8) | 37.02977 | 8 | *0.00001136 |
| EI | 18F&G | 9 (9 & 7) | 16.77821 | 9 | 0.05230469 |
| EI | 19F&G | 10 (4&10) | 26.57795 | 10 | 0.00303591 |
| VE | 18&19F | 10 (10&10) | 24.89894 | 10 | 0.00554038 |
| VE | 18&19G | 10 (6&10) | 18.16196 | 10 | 0.05229223 |

* Chi-Square significant at alpha = .05/40 = .00125

Table 13

Sample 1: Grade 11
Likelihood Ratio Chi-Square Tests of Significance Item Order Effects

| Subtest | Form | Degree of Polynomial | Chi-Square | D.F. | Probability |
|---------|--------|----------------------|------------|------|-------------|
| GS | 18F&G | 7 (7& 6) | 24.17904 | 7 | *4.00105965 |
| GS | 19F&G | 10 (4&10) | 21.67304 | 10 | 0.01685969 |
| GS | 18F&G | 10 (6&10) | 13.73382 | 10 | 0.18548293 |
| GS | 19F&G | 7 (7 & 7) | 7.36182 | 7 | 0.39220368 |
| AR | 18&19F | 10 (7&10) | 20.46122 | 10 | 0.02518006 |
| AR | 18&19G | 10 (7&10) | 7.62054 | 10 | 0.66584670 |
| WK | 18&19F | 10 (10& 7) | 23.46679 | 10 | 0.00914872 |
| WK | 18&19G | 10 (10&10) | 16.45481 | 10 | 0.08733210 |
| PC | 18&19F | 8 (8 & 8) | 325.10830 | 8 | *0.00000000 |
| PC | 18&19G | 8 (5 & 8) | 105.73147 | 8 | *0.00000000 |
| CS | 18F&G | 9 (9& 8) | 28.41594 | 9 | *0.00081255 |
| CS | 19F&G | 8 (7 & 8) | 17.79410 | 8 | 0.02282426 |
| AS | 18F&G | 10 (8&10) | 24.28925 | 10 | 0.00686855 |
| AS | 19F&G | 9 (7 & 9) | 7.40078 | 9 | 0.59546732 |
| MK | 18&19F | 10 (8&10) | 12.48965 | 10 | 0.25362133 |
| MK | 18&19G | 10 (10& 8) | 18.04874 | 10 | 0.05414699 |
| MC | 18F&G | 10 (10& 7) | 6.85449 | 10 | 0.73910647 |
| MC | 19F&G | 10 (9&10) | 23.97514 | 10 | 0.00766666 |
| EI | 18F&G | 10 (10& 7) | 3.63330 | 10 | 0.96237740 |
| EI | 19F&G | 10 (9&10) | 28.48896 | 10 | 0.00150673 |
| VE | 18&19F | 10 (9&10) | 37.81613 | 10 | *0.00004085 |
| VE | 18&19G | 9 (9 & 8) | 17.12335 | 9 | 0.04681908 |

* Chi-Square significant at alpha = .05/40 = .00125

Table 14a
Sample 2: Grade 12
STD-AFQT IOT&E METRIC ANOVA

| <u>Source</u> | <u>DF</u> | <u>Type III SS</u> | <u>% of Total Model Variance¹</u> | <u>Mean Square</u> | <u>F Value</u> | <u>Pr > F</u> |
|----------------|-----------|--------------------|--|--------------------|----------------|------------------|
| MEPSNO | 67 | 5330646.37340118 | 93.94% | 79561.88617017 | 112.54 | 0.0001 |
| FORM | 4 | 64104.03977559 | 1.13% | 16026.00994390 | 22.67 | 0.0001 |
| MEPSNO BY FORM | 268 | 280028.43967705 | 4.93% | 1044.88223760 | 1.48 | 0.0001 |
| Total Model | 339 | 5674778.8528538 | 100.00% | | | |

Table 14b
Sample 2: Grade 12
STD-AFQT IOT&E METRIC ANOVA
LEAST SQUARES MEANS

| <u>Form</u> | <u>LSMEAN</u> |
|-------------|---------------|
| REF | 201.131772 |
| 18F | 202.580004 |
| 18G | 201.346864 |
| 19F | 203.141072 |
| 19G | 201.223384 |

¹ Here, the "total model variance" means the sum of the SAS Type III SS over the effects listed.

Table 15a

**Sample 2: Grade 11
STD-AFQT IOT&E METRIC ANOVA**

| <u>Source</u> | <u>DF</u> | <u>Type III SS</u> | <u>% of Total Model Variance²</u> | <u>Mean Square</u> | <u>F Value</u> | <u>Pr > F</u> |
|-------------------|-----------|--------------------|--|--------------------|----------------|------------------|
| MEPSNO | 67 | 8270975.60889049 | 95.82% | 123447.39714762 | 177.22 | 0.0001 |
| FORM | 4 | 93921.81254488 | 1.09% | 23480.45313622 | 33.71 | 0.0001 |
| MEPSNO BY FORM | 268 | 267029.99571112 | 3.09% | 996.38058101 | 1.43 | 0.0001 |
| Total Model | 339 | 8631927.4171465 | 100.00% | | | |

Table 15b

**Sample 2: Grade 11
STD-AFQT IOT&E METRIC ANOVA
LEAST SQUARES MEANS**

| <u>Form</u> | <u>LSMEAN</u> |
|-------------|---------------|
| REF | 201.863924 |
| 18F | 203.235208 |
| 18G | 201.883637 |
| 19F | 203.157370 |
| 19G | 201.458092 |

² Here, the "total model variance" means the sum of the SAS Type III SS over the effects listed.

Table 16a
Sample 2: Grade 12
STD-AFQT RE-EQUATED METRIC ANOVA

| <u>Source</u> | <u>DF</u> | <u>Type III SS</u> | <u>% of Total Model Variance³</u> | <u>Mean Square</u> | <u>F Value</u> | <u>Pr > F</u> |
|----------------|-----------|--------------------|--|--------------------|----------------|------------------|
| MEPSNO | 67 | 5498255.30862095 | 94.85% | 82063.51206897 | 113.23 | 0.0001 |
| FORM | 4 | 11035.38766876 | 0.19% | 2758.84691719 | 3.81 | 0.0043 |
| MEPSNO BY FORM | 268 | 287716.93913661 | 4.96% | 1073.57066842 | 1.48 | 0.0001 |
| Total Model | 339 | 5797007.6354263 | 100.00% | | | |

Table 16b
Sample 2: Grade 12
STD-AFQT RE-EQUATED METRIC ANOVA
LEAST SQUARES MEANS

| <u>Form</u> | <u>LSMEAN</u> |
|-------------|---------------|
| REF | 201.131772 |
| 18F | 201.155770 |
| 18G | 201.866160 |
| 19F | 201.058747 |
| 19G | 201.727998 |

³ Here, the "total model variance" means the sum of the SAS Type III SS over the effects listed.

Table 17a
Sample 2: Grade 11
STD-AFQT RE-EQUATED METRIC ANOVA

| <u>Source</u> | <u>DF</u> | <u>Type III SS</u> | <u>% of Total Model Variance⁴</u> | <u>Mean Square</u> | <u>F Value</u> | <u>Pr > F</u> |
|----------------|-----------|--------------------|--|--------------------|----------------|------------------|
| MEPSNO | 67 | 8532821.11657319 | 96.84% | 127355.53905333 | 176.92 | 0.0001 |
| FORM | 4 | 4576.74969837 | 0.05% | 1144.18742459 | 1.59 | 0.1740 |
| MEPSNO BY FORM | 268 | 274049.93214366 | 3.11% | 1022.57437367 | 1.42 | 0.0001 |
| Total Model | 339 | 8811447.7984152 | 100.00% | | | |

Table 17b
Sample 2: Grade 11
STD-AFQT RE-EQUATED METRIC ANOVA
LEAST SQUARES MEANS

| <u>Form</u> | <u>LSMEAN</u> |
|-------------|---------------|
| REF | 201.864027 |
| 18F | 201.773236 |
| 18G | 201.959851 |
| 19F | 201.636314 |
| 19G | 201.502001 |

⁴ Here, the "total model variance" means the sum of the SAS Type III SS over the effects listed.

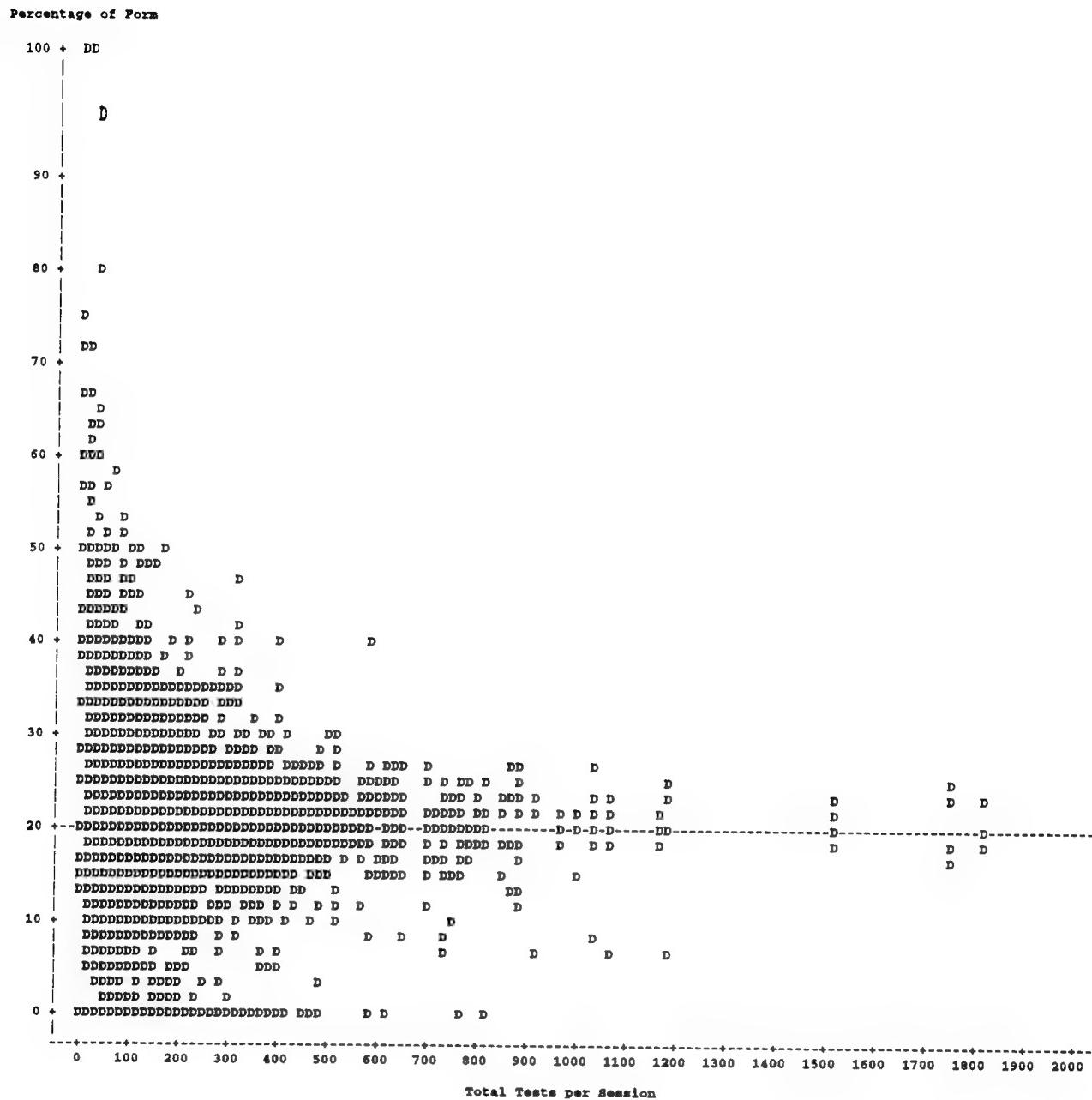


Figure 1a. Percentage of Form (18F, 18G, 19F, 19G, & Reference) Administered in a "Deleted" Session by Total Number of Forms Administered Per Session for Sample #1.

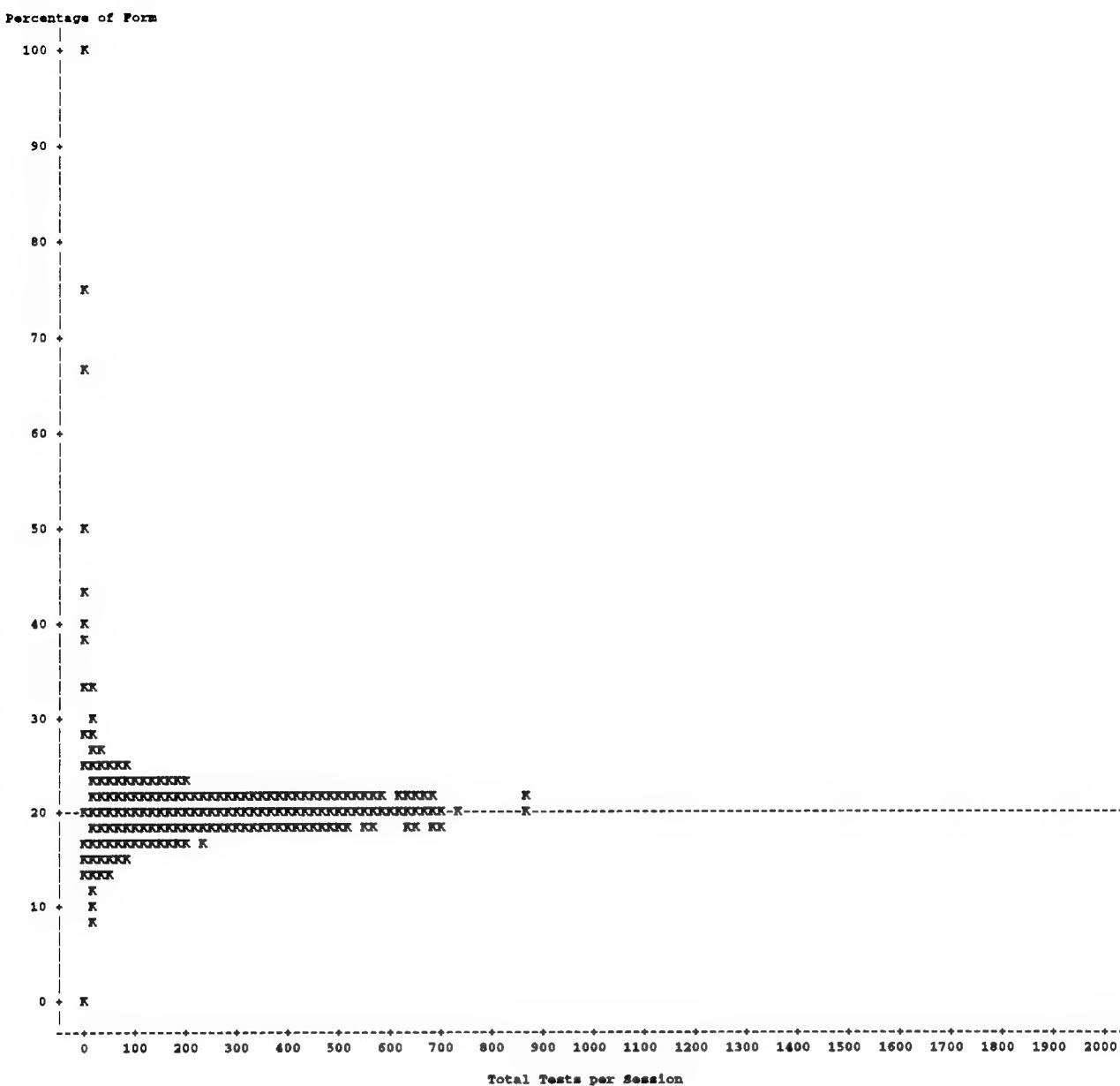


Figure 1b. Percentage of Form (18F, 18G, 19F, 19G, & Reference) Administered in a "Kept" Session by Total Number of Forms Administered Per Session for Sample #1.

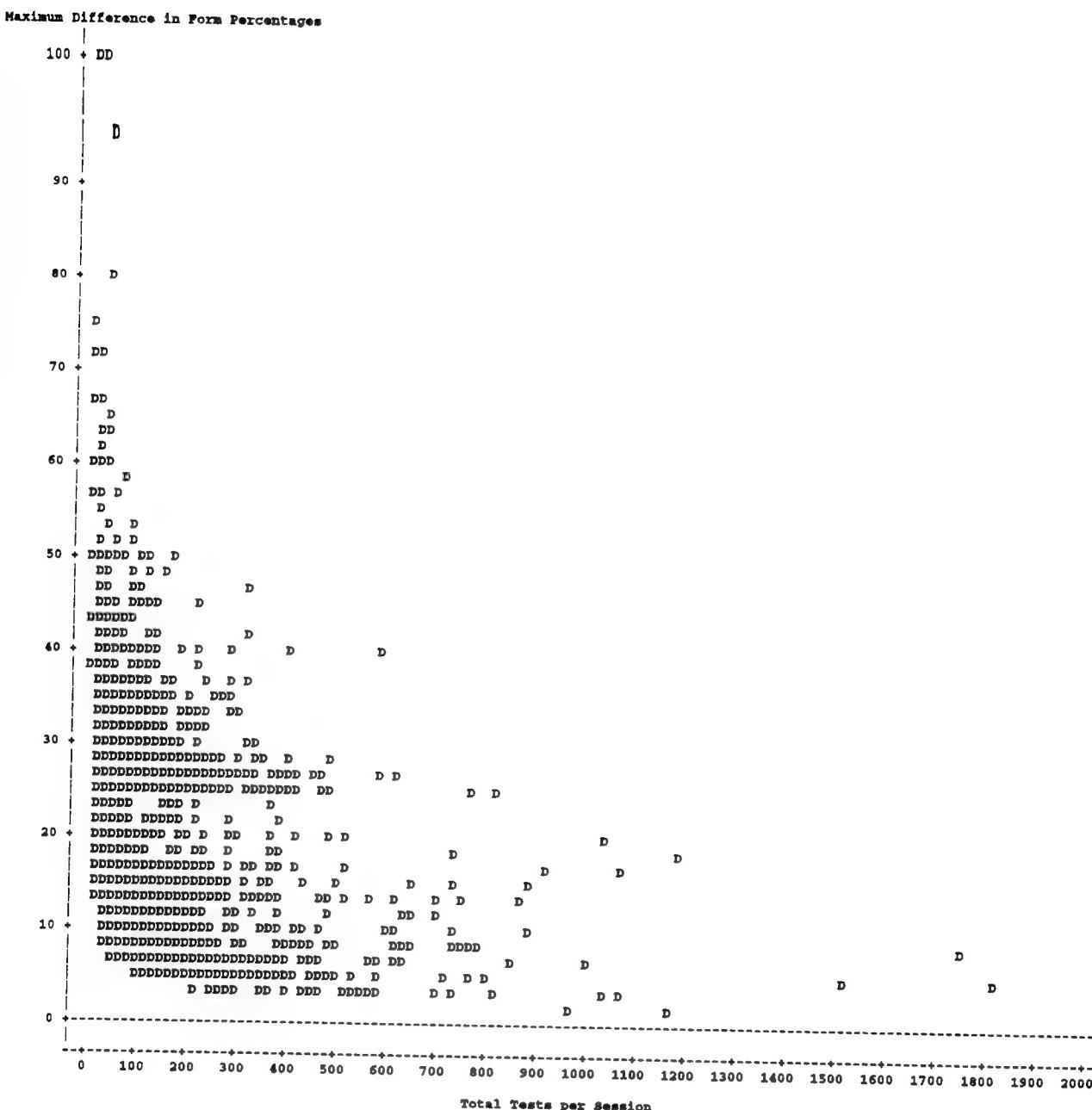


Figure 1c. Percentage Difference (Most Administered Form -Least Administered Form) in a "Deleted" Session by Total Number of Forms Administered Per Session for Sample #1.

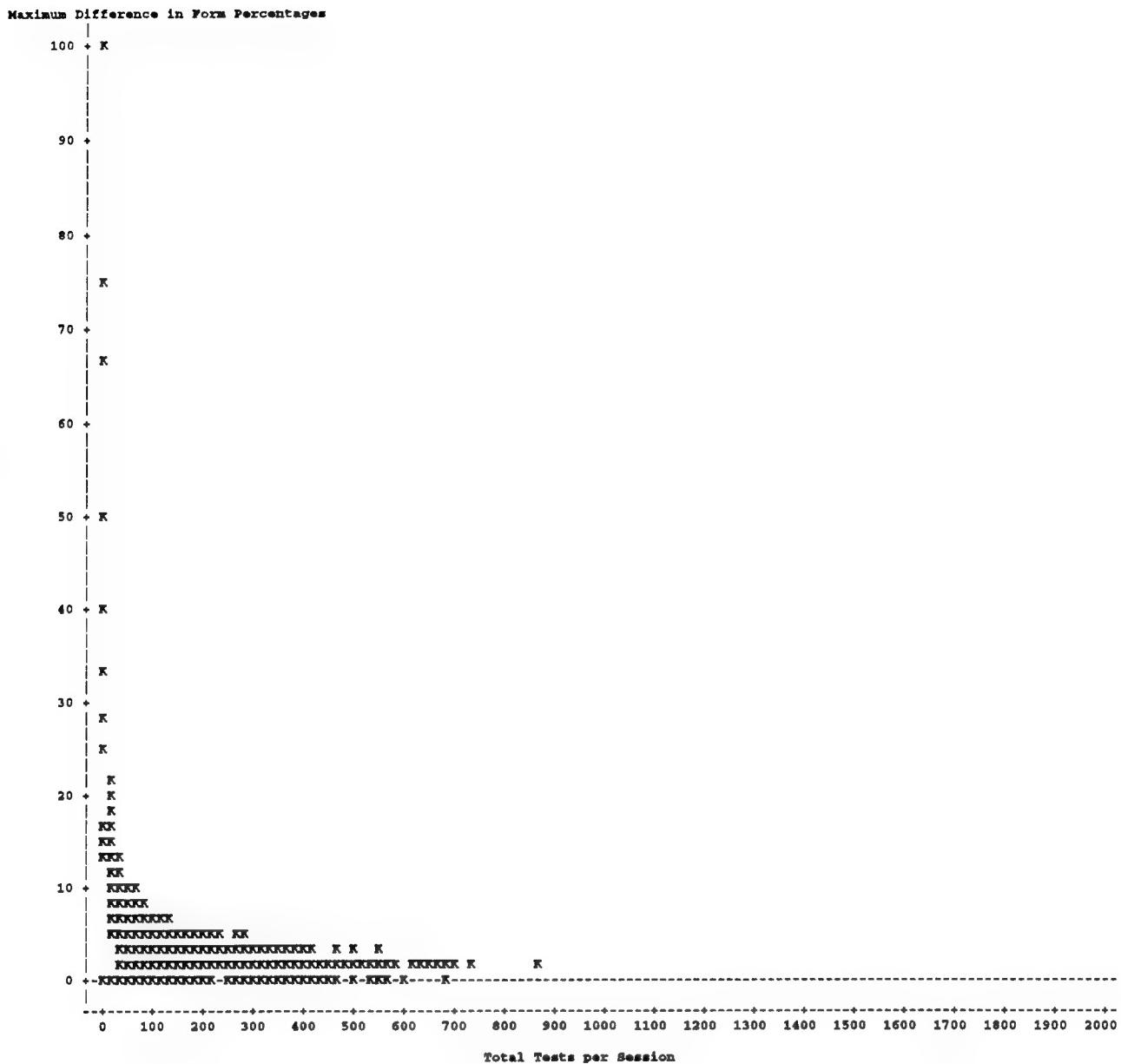


Figure 1d. Percentage Difference (Most Administered Form -Least Administered Form) in a "Kept" Session by Total Number of Forms Administered Per Session for Sample #1.

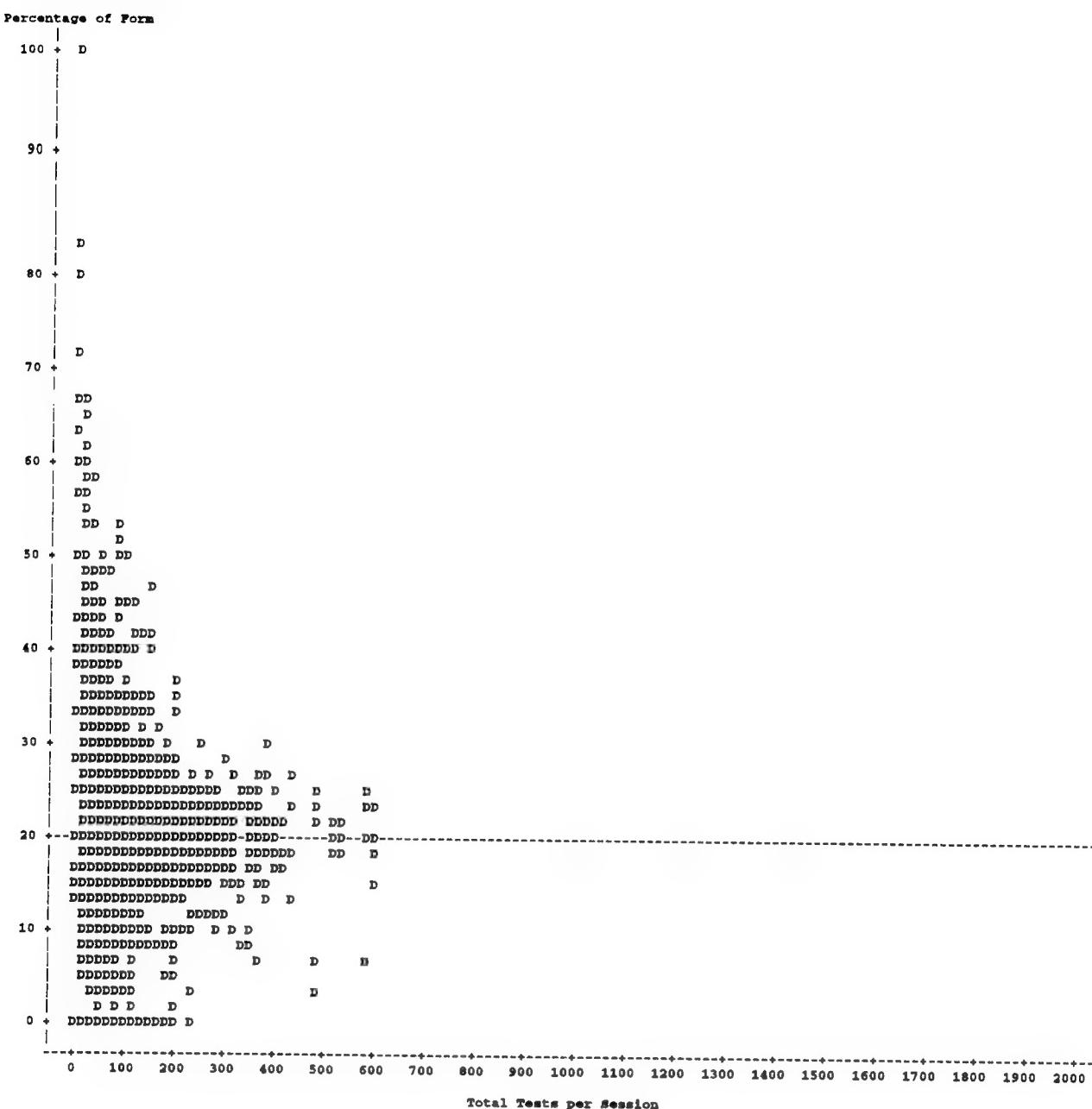


Figure 2a. Percentage of Form (18F, 18G, 19F, 19G, & Reference) Administered in a "Deleted" Session by Total Number of Forms Administered Per Session for Sample #2

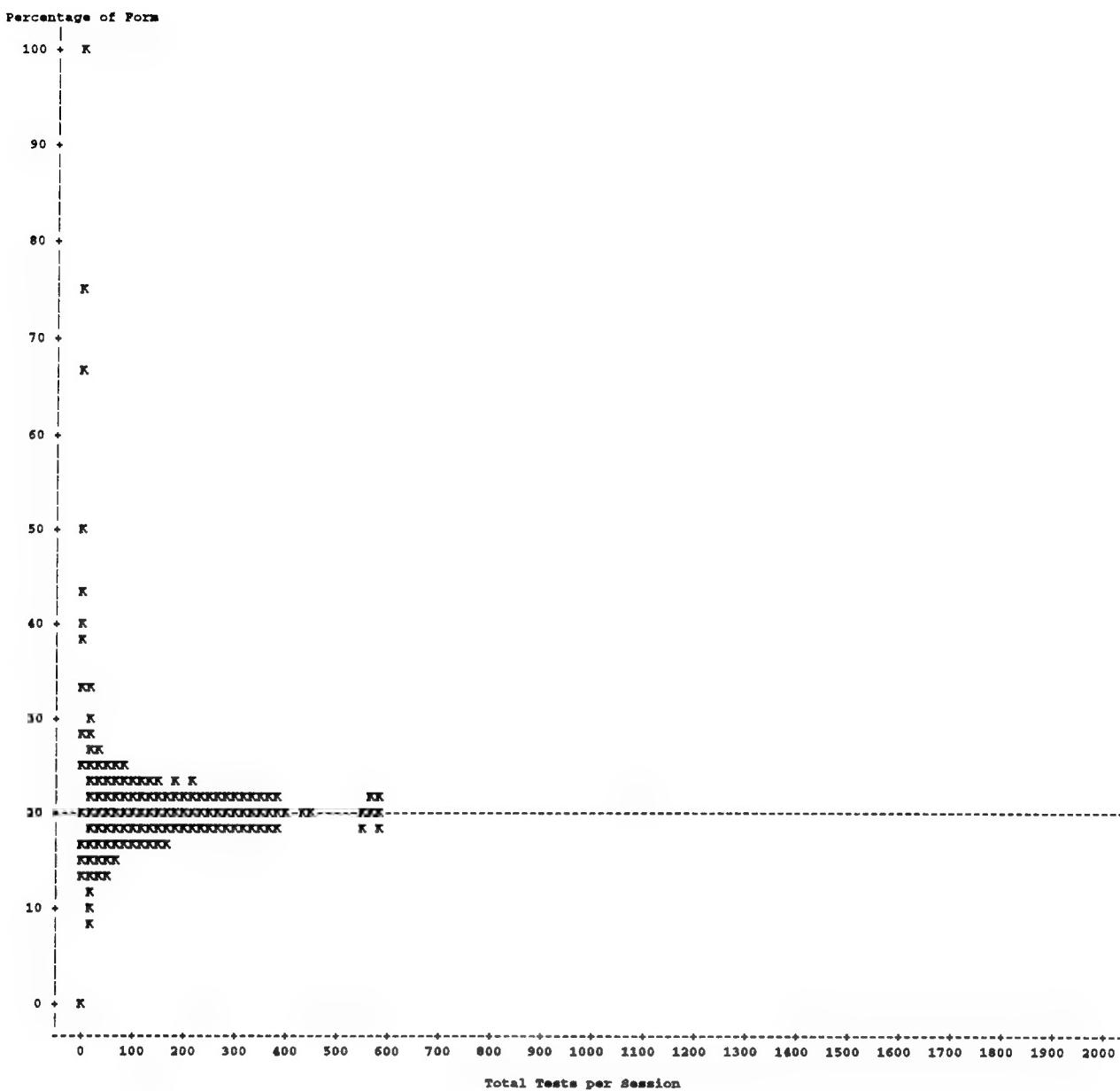


Figure 2b. Percentage of Form (18F, 18G, 19F, 19G, & Reference) Administered in a "Kept" Session by Total Number of Forms Administered Per Session for Sample #2.

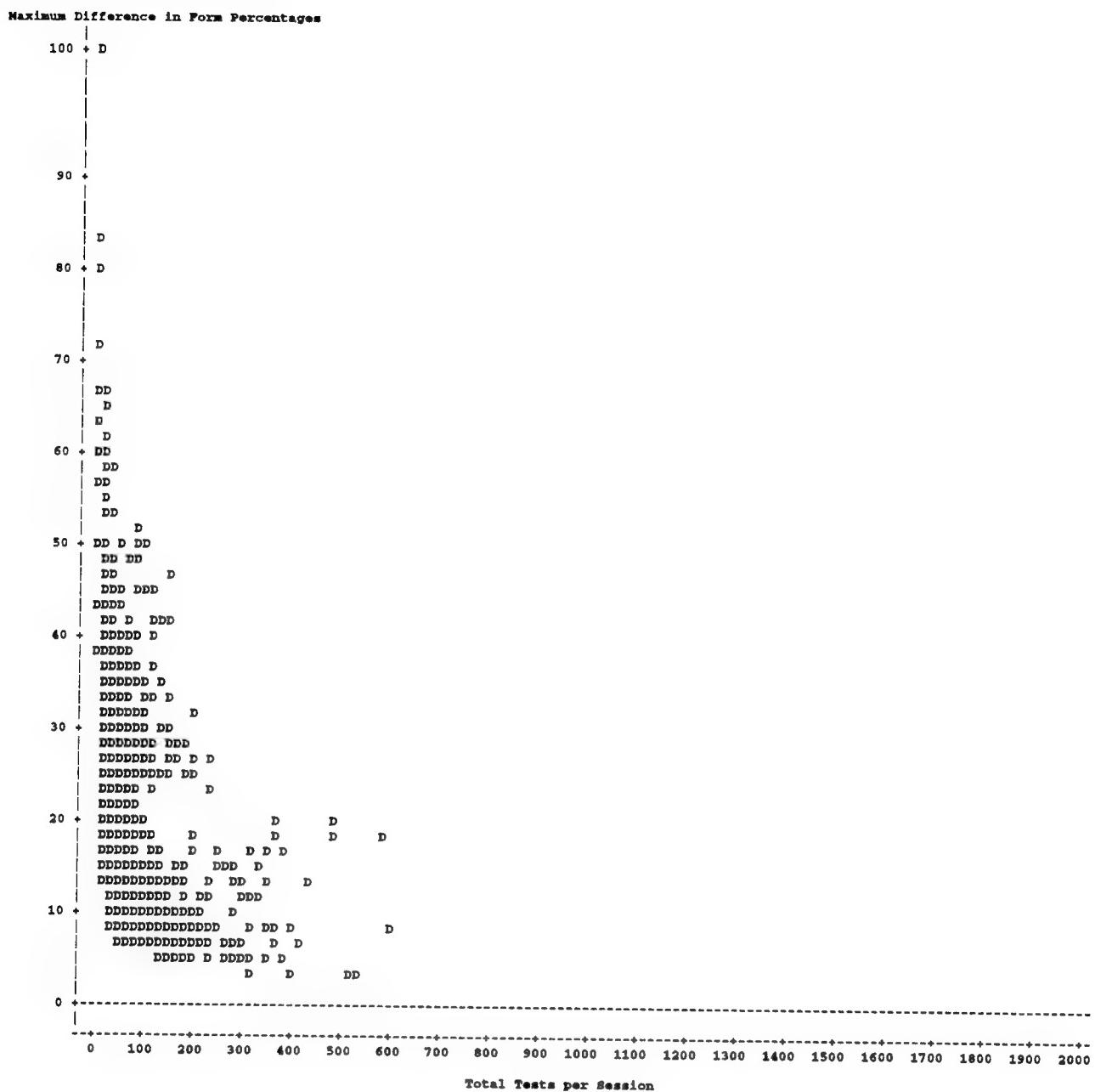


Figure 2c. Percentage Difference (Most Administered Form -Least Administered Form) in a "Deleted" Session by Total Number of Forms Administered Per Session for Sample #2.

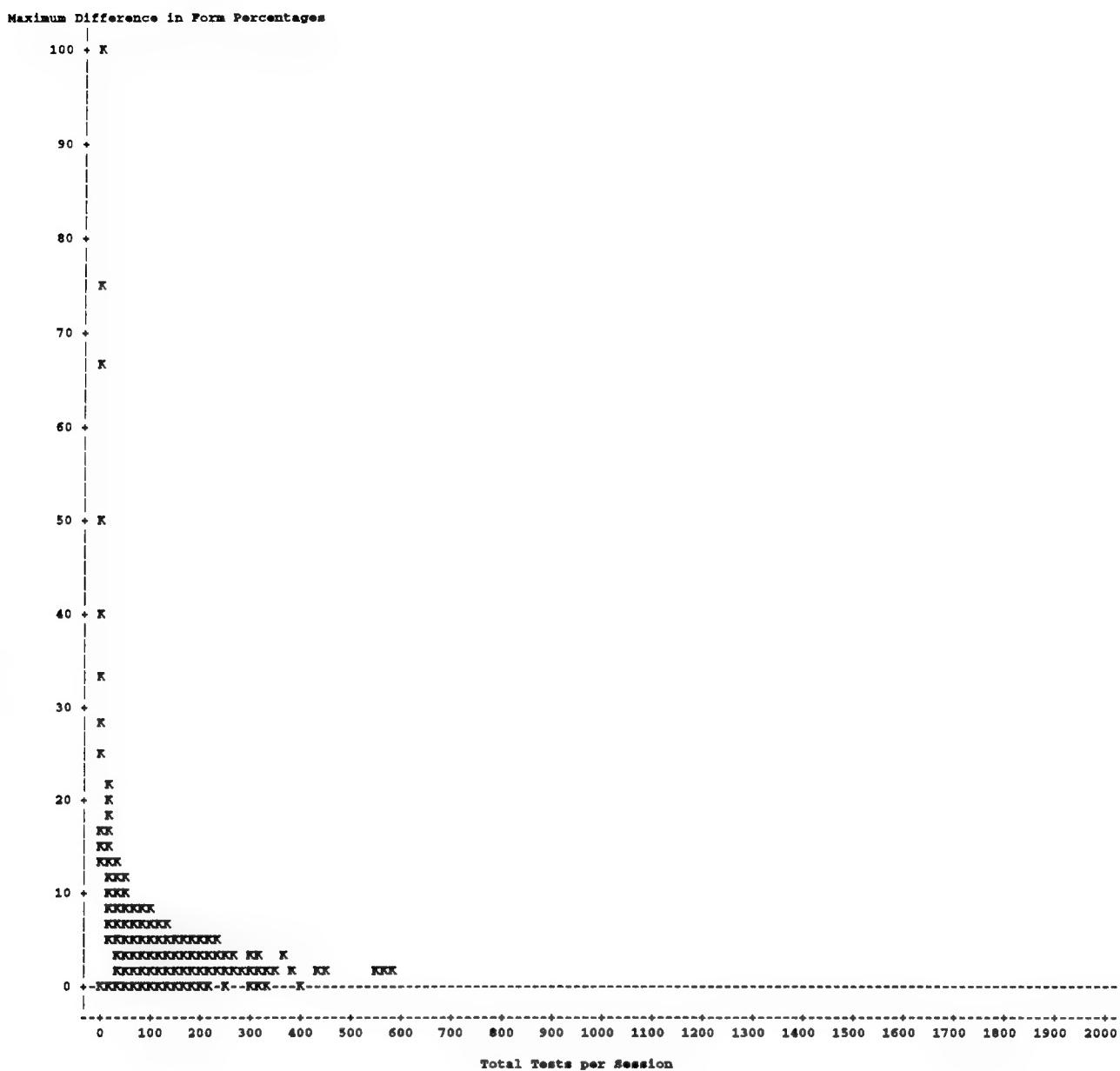


Figure 2d. Percentage Difference (Most Administered Form -Least Administered Form) in a "Kept" Session by Total Number of Forms Administered Per Session for Sample #2.

**Form 18f AFQT
Grade 12 Sample 2**

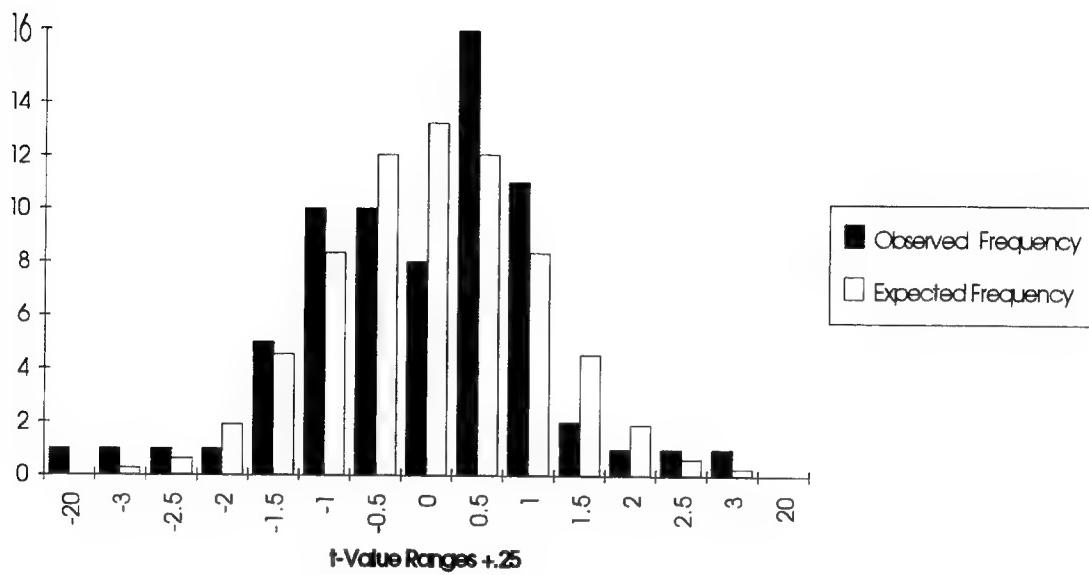


Figure 3. Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 18F versus Reference Form: Grade 12; Sample #2.

**Form 18f AFQT
Grade 11 Sample 2**

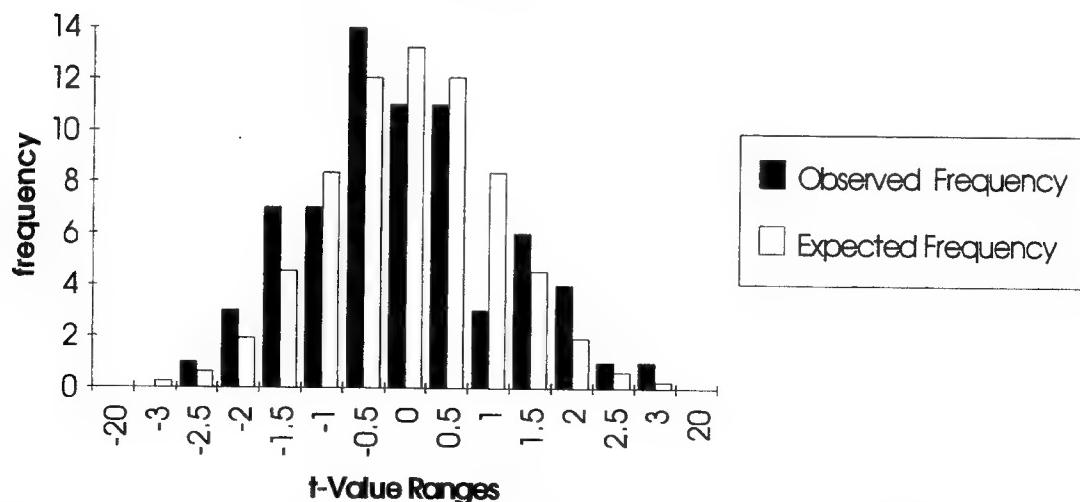


Figure 4. Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 18F versus Reference Form: Grade 11; Sample #2.

**Form 19f AFQT
Grade 12 Sample 2**

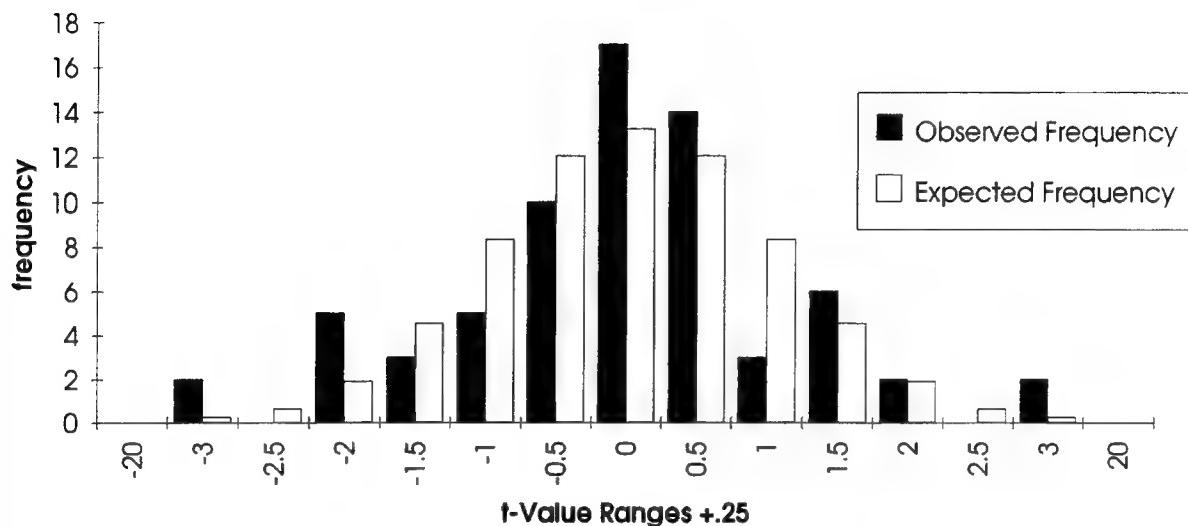


Figure 5. Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 19F versus Reference Form; Grade 12; Sample #2.

**Form 19f AFQT
Grade 11 Sample 2**

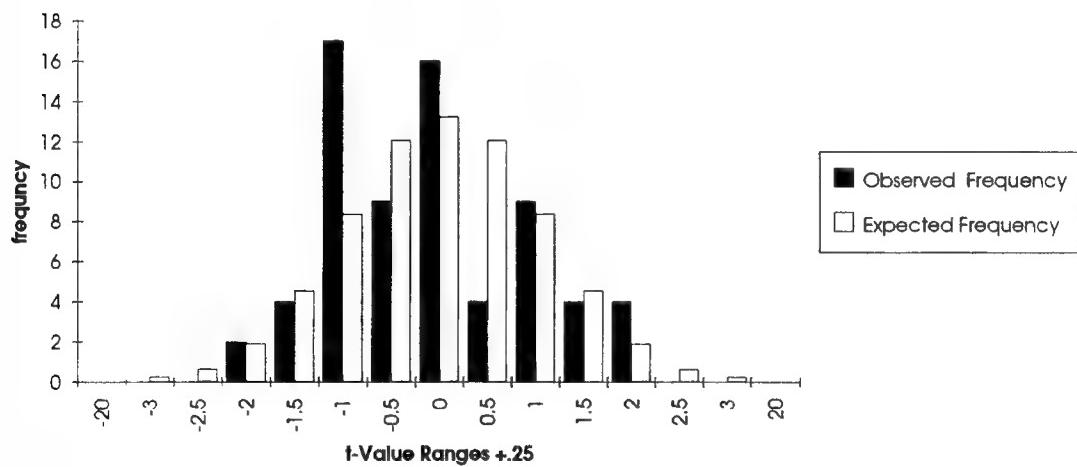


Figure 6. Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 19F versus Reference Form; Grade 11; Sample #2

**Form 18g AFQT
Grade 12 Sample 2**

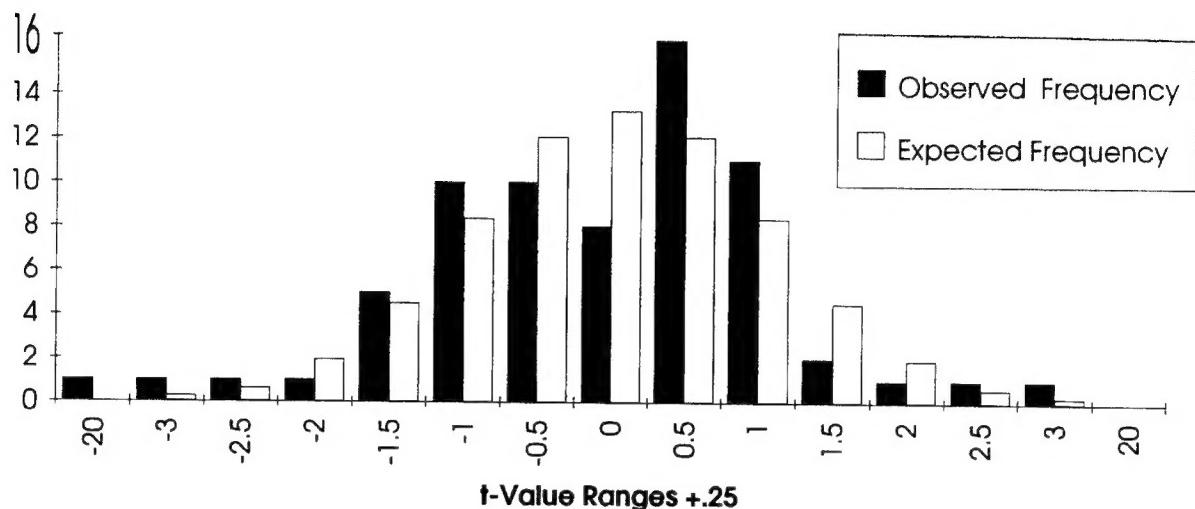


Figure 7. Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 18G versus Reference Form: Grade 12; Sample #2.

**Form 18g AFQT
Grade 11 Sample 2**

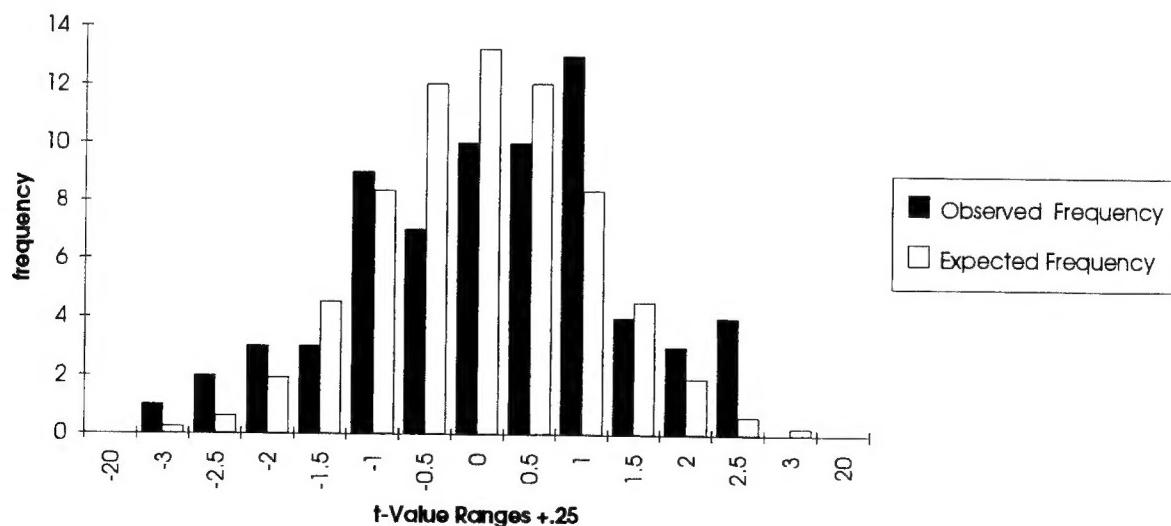


Figure 8. Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 18G versus Reference Form: Grade 11; Sample #2.

**Form 19g AFQT
Grade 12 Sample 2**

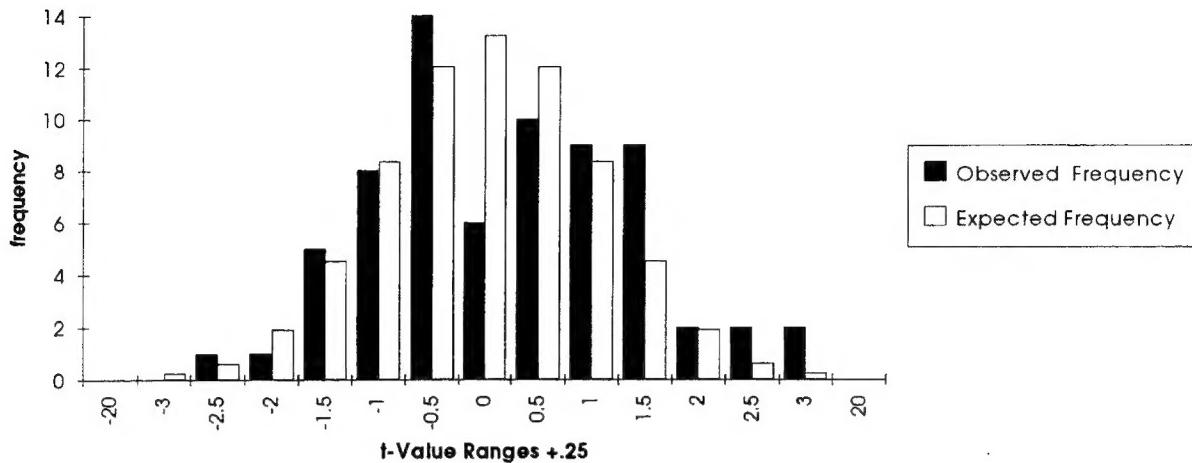


Figure 9. Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 19G versus Reference Form: Grade 12; Sample #2.

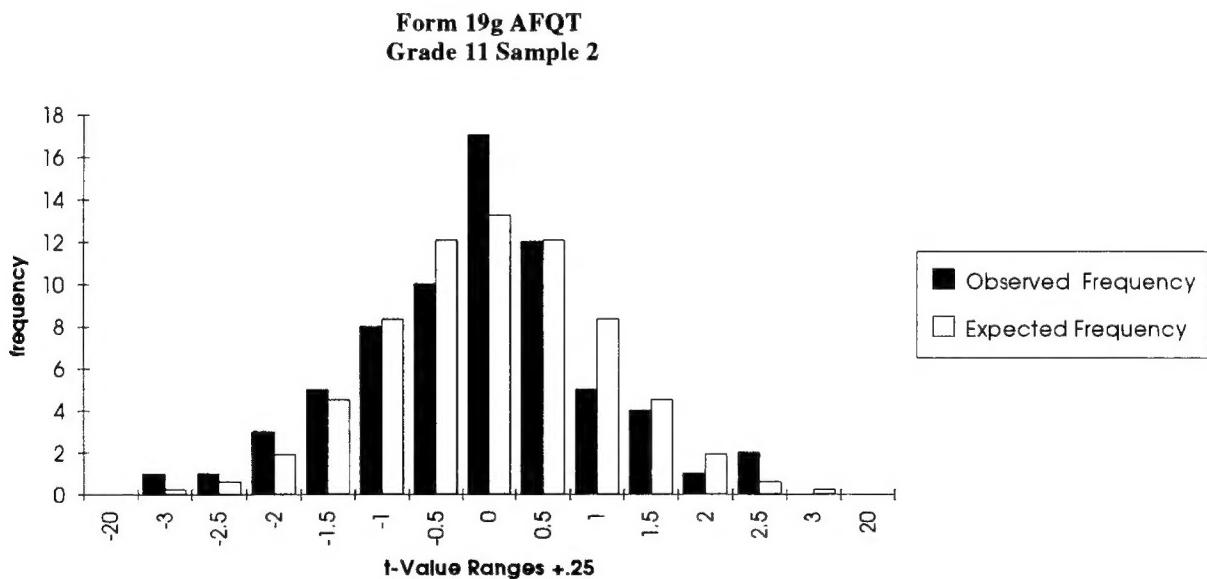


Figure 10. Histogram of T-Test Statistics for Mean Difference of Re-equated STD-AFQT Score: Form 19G versus Reference Form: Grade 11; Sample #2.

Form 18f AFQT
Grade 11 v 12 Sample 2

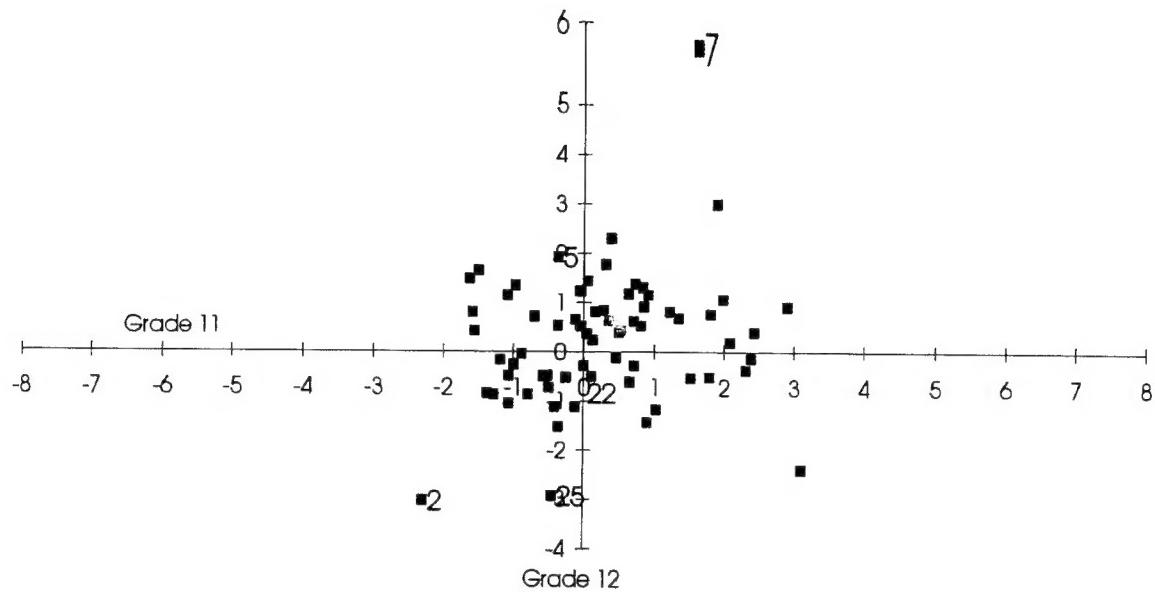


Figure 11. Bivariate Plot of Grade 12 by Grade 11 T-Test Statistics for Mean Difference of STD-AFQT Score: Form 18F versus Reference Form: Sample #2.

Form 19f AFQT
Grade 11 v 12 Sample 2

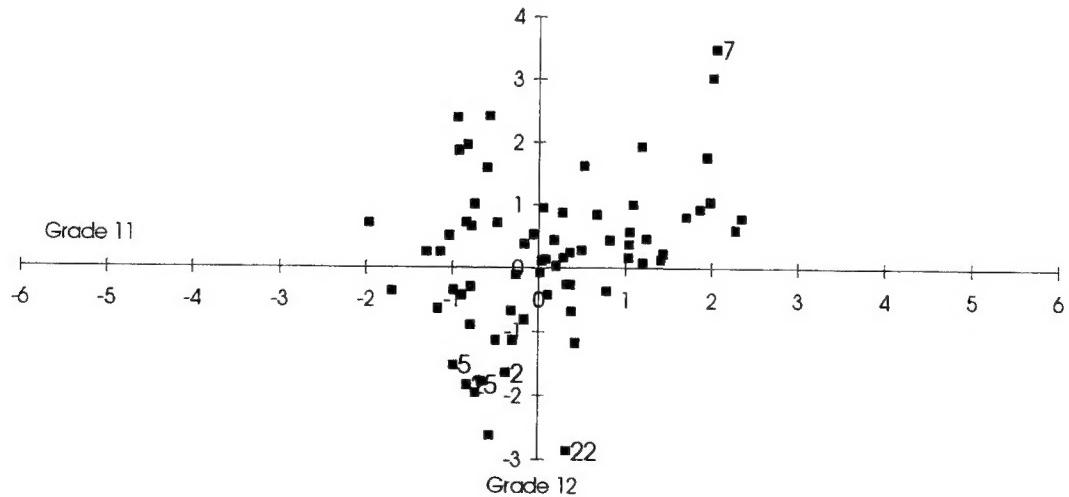


Figure 12. Bivariate Plot of Grade 12 by Grade 11 T-Test Statistics for Mean Difference of STD-AFQT Score: Form 19F versus Reference Form: Sample #2.

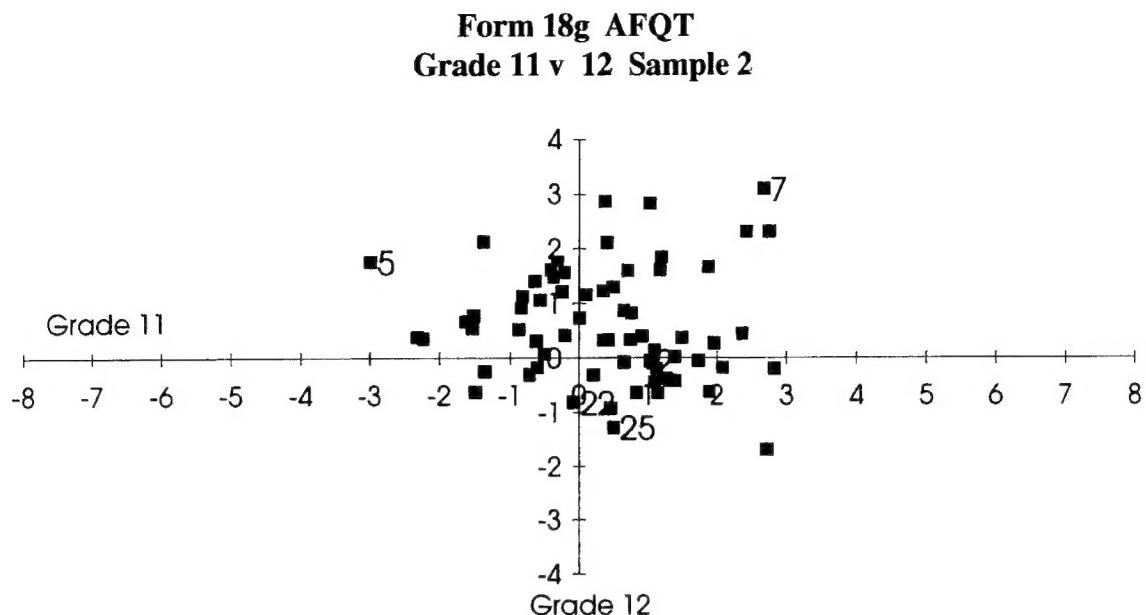


Figure 13. Bivariate Plot of Grade 12 by Grade 11 T-Test Statistics for Mean Difference of STD-AFQT Score: Form 18G versus Reference Form: Sample #2.

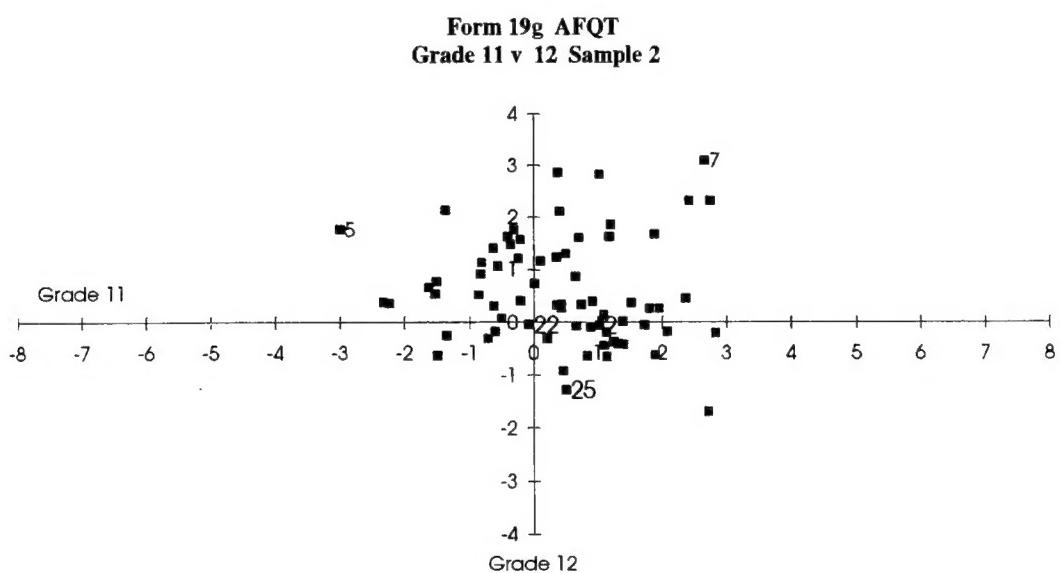


Figure 14. Bivariate Plot of Grade 12 by Grade 11 T-Test Statistics for Mean Difference of STD-AFQT Score: Form 19G versus Reference Form: Sample #2.